UNIVERSITY OF CANTERBURY

DISTRIBUTION OF THE LIKELIHOOD RATIO IN FERMI GAMMA RAY MODELS

by

Benjamin Roberts

A thesis submitted in partial fulfillment for the
degree of Master of Science

in the

Department of Physics and Astronomy
College of Science

September 2017

UNIVERSITY OF CANTERBURY

Abstract

Department of Physics and Astronomy
College of Science

Master of Science

by Benjamin Roberts

Gamma rays, the highest frequency component of the electromagnetic spectrum, are produced by the highest energy objects in the universe. The list of sources includes pulsars, active galactic nuclei, cosmic ray interactions, and as some have suggested dark matter. The Fermi Large Area Telescope was launched into orbit in 2008 to collect all-sky gamma ray data to investigate the mechanisms and sources that generate these particles.

Models are fit to the data through the use of the maximum likelihood method, in which a statistical algorithm finds the most likely model parameters that generated the data. When comparing two models to the same data, the likelihood ratio is developed that gives an indication as to how well the models explain the data.

In some problems, particularly in the detection of point sources against a background, it is of interest to compare the best fitting model to data with and without a point source present. In this case, the likelihood ratio that is used will determine whether there is a statistically significant source present.

This study will motivate the use of the test statistic and examine the asymptotic distribution of the statistic when comparing models. This will then be extended to consider a multiband analysis of data, and how the test statistic used is distributed in these cases. Using Monte Carlo techniques, the asymptotic distributions of the test statistic will be validated. This will be achieved within the FermiTools suite using the Fermi Science Support Center recommended procedures in fitting models to gamma ray data.

Acknowledgements

I would like to thank my supervisor Dr Chris Gordon of the University of Canterbury. His support during the previous year of this project has been very much appreciated in offering guidance, direction, help, and a sounding board for my work. His office door was always open whenever I required his input.

Acknowledgement is also given to Dr Oscar Macias of Virginia Tech for the many emails and Skype discussions on the technical aspects of FermiTools and problem solving throughout this project. Without your generous help, I would have had trouble overcoming some of the hurdles in this research.

Finally, I must express my gratitude to my parents and family for their continuous encouragement throughout my years of study. To my partner, Ally, for your help in keeping me sane in rough patches of this research and supporting me during the year. Adam, Jonty, Ryan, and Liam, friends who were going through the trials of postgraduate physics as well, thank you for the support and comradery.

This accomplishment would not have been possible without you all. Thank you.

Contents

List of Figures

Abbreviations

ACD	**A**nti-**C**oincidence **D**etector
AGN	**A**ctive **G**alactic **N**uclei
DAQ	**D**ata **Acq**uisition System
EGRET	**E**nergetic **G**amma **R**ay **E**xpirement **T**elescope
FSSC	**F**ermi **S**cience **S**upport **C**enter
FITS	**F**lexible **I**mage **T**ransport **S**ystem
GBM	**G**amma Ray **B**urst **M**onitor
GTI	**G**ood **T**ime **I**nterval
HE	**H**igh **E**nergy
IRF	**I**nstrument **R**esponse **F**unction
LAT	**L**arge **A**rea **T**elescope
MLE	**M**aximum **L**ikelihood **E**stimator
PSF	**P**oint **S**pread **F**unction
ROI	**R**egion **o**f **I**nterest
SC	**S**pace **C**raft
SNR	**S**uper**n**ovae **R**emnant
SR	**S**ource **R**egion
SSD	**S**ilicon **S**trip **D**etector
TKR	**T**rac**k**e**r**

For Mum and Dad,
who always supported and nurtured my love of the sciences

Chapter 1

Overview

Astrophysical sources of gamma rays are some of the most violent physical processes known, and often represent the extreme of high energy physics. The Fermi Gamma Ray Space Telescope was launched in 2008 to study these sources, and in the (continuing) nine year run of data collection it has provided the most resolved all-sky gamma ray flux map since the origin of gamma ray astronomy in the 1950's.

In Chapter Two, the field of gamma ray astronomy will be introduced. In particular this will detail the production and sources of gamma rays, and their interactions with matter. These interactions motivate a number of constraints on how telescopes must be designed to be effective at collecting gamma ray data. It will also be shown how these constraints have been incorporated into the design of the Fermi Large Area Telescope.

As one can not distinguish the photons detected from one source rather than another with ease, due to the relatively large angular resolution, a suitable method of fitting models must be developed. This will be discussed in Chapter Three, as models of gamma ray sources are described and folded with the Fermi Instrument Response Functions to create an expected number of counts from the model. This quantity can thus be compared between model and data.

Many photon counting experiments, such as Fermi gamma ray data, utilise the maximum likelihood method in model fitting. The likelihood function measures the probability of obtaining data given a particular model, and by maximising this optimisation problem one can infer parameter estimates and their uncertainties. This process will be expanded upon in Chapter Four.

The FermiTools package, created for the study of Fermi data, provide a series of functions that allow the maximum likelihood method to be achieved practically. In Chapter Five, the process of simulating or importing collected data and the creation of necessary

model-independent quantities with these functions will be discussed. These files will be used as inputs to the `gtlike` function that optimises the model parameters to maximise the likelihood function.

Model-independent, or multiband, fits allow a study to use motivated assumptions and fit simple models independently in multiple small energy bins. This is opposed to a more general method of committing to a particular model and fitting it across an entire energy range of interest. Chapter Six will introduce the likelihood ratio, a test statistic to measure the goodness of fit of a model, and derive the asymptotic distribution of this statistic for band by band fits.

This study will present an investigation into these multiband analyses using simulated gamma ray data and the FermiTools suite to perform model fitting. A uniform background source is simulated for typical energy ranges and fluxes seen in Fermi studies, and fit both with and without a point source present. If there are statistically significant deviations from the base model, the likelihood ratio will indicate this. This problem is similar to cases of source detection. This process will be repeated 1000 times to create a Monte Carlo sample. It will be shown that the empirical likelihood ratio distribution will indeed follow the theoretical distribution derived in the previous chapter.

It has been suggested that the pair annihilation or decay of dark matter particles could be an observable source of gamma rays near the Galactic Center. Due to the density of gas fields and numerous other sources, the signal is difficult to study and significant uncertainties affect the interpretation of any results. Detailed studies of this region including careful modelling of diffuse and point sources, may yield more fruitful evidence of dark matter. It is hoped that the distributions derived and shown in this study can be used in a multiband analysis of the Galactic Center region at a later date.

Chapter 2

Gamma Ray Astronomy

The gamma ray band of the electromagnetic spectrum was discovered in 1900 in a study of the radiation emitted by radium isotopes[13]. By the 1950's a number of scientists had made predictions that, among other processes, pion decay, bremsstrahlung processes, and synchrotron radiation from strong magnetic fields would occur in astrophysical sources. However at the time, detectors were not able to measure the astrophysical gamma ray spectrum.

The SAS-2 telescope launched in 1972[14], followed by the COS-B telescope of 1975[15], provided the first detailed sky maps in the gamma ray spectrum. The data allowed the discovery of a number of bright sources and pioneered the research in this discipline. Following the success of these missions, and the developments in detection technologies, the Compton Gamma Ray Observatory was designed by NASA and launched in 1991[16]. Carrying EGRET, this mission discovered nearly 300 sources and was able to study and classify many of their spectra.

The Fermi Gamma Ray Space Telescope, carrying the Large Area Telescope, was intended as the successor to EGRET following the end of that mission. In this chapter, the field of gamma ray astronomy will be introduced describing the production of gamma rays and their interaction with matter, motivating the design of the Fermi telescope. The various astrophysical sources of gamma rays will be discussed and the characteristics of their spectra.

Throughout this project, and generally throughout the literature, the gamma ray band of the electromagnetic spectrum is taken to extend in energy from soft gamma rays of 100 keV through to hard gamma rays of 100 TeV. Of interest to the data collected by the Fermi telescope are the 'GeV gamma rays' or the high energy band, which is understood as 100 MeV to 100's of GeV.

2.1 Gamma Ray Astronomy

The production of gamma rays are limited due to their high energy, such that they are not generally produced in significant quantities persistently. Consider a blackbody radiator that peaks at say 1 MeV in the high energy band, the radiator would have a temperature in the order of 10^{10} K. This can only be achieved in the most extreme of scenarios, such as core collapse.

However, non-thermal gamma rays can be produced by persistent low flux sources. Through processes such as particle decays, and the electromagnetic interactions of charged particles. These will be discussed briefly.

While a number of decay processes produce high energy photons, the neutral pion (π^0) decay produces significant contributions to the flux of a gamma ray source[17]. This arises in astrophysical sources when cosmic ray protons interact with the gas and dust of the interstellar medium, producing a pair of gamma rays. The energy spectrum of this peaks strongly at 67.5 MeV, half the π^0 mass, falling off at a power law rate.

It is speculated that in some particle dark matter theories the decay or pair annihilation of dark matter particles will produce gamma rays in some quantity[18]. However, the processes themselves are at best speculative and the evidence for their gamma ray signal requires further study and more detailed data analysis[19][20].

The other side of non-thermal gamma ray production is the electromagnetic interaction with charged particles. This occurs in three ways, namely bremsstrahlung and synchrotron radiation, and Inverse Compton scattering [17].

When charged particles interact in a gas or dust field, bremsstrahlung radiation is produced. Similarly, synchrotron radiation is produced in the interaction with a magnetic field. The spectra of these two interactions depend strongly on the spectra of the population of charged particles. Inverse Compton scattering occurs in the interaction of soft photon radiation fields with charged leptons, with a spectra depending on the target radiation field. In some instances the radiation field can be provided by the photons produced by the synchrotron processes, this is referred to as synchrotron self-Compton scattering.

Abundant in the Milky Way galaxy is an isotropic population of charged cosmic rays. These particles generally provide the input to create the above defined gamma rays[21]. The origin of cosmic rays are primarily thought to come from supernova remnants[1]. During the final phases of a massive stars evolution, the star explodes in a dramatic way expelling the majority of the stars material at high speed. Ambient matter surrounding the supernova forms a shock wave emanating away from the source. The vast energies of

the shockwave create a shell of plasma, and in a process called Fermi acceleration create high speed charged particles[22][23]. At the front of the plasma shell, charged particles can cross the shockwave and be reflected back. This occurs many times, crossing back and forth, and gaining energy with each pass. After many passes, the particle has sufficient energy to escape the expanding shock wave, escaping as cosmic rays.

It is believed that similar processes accelerate particles in extragalactic objects, beyond the Milky Way[24].

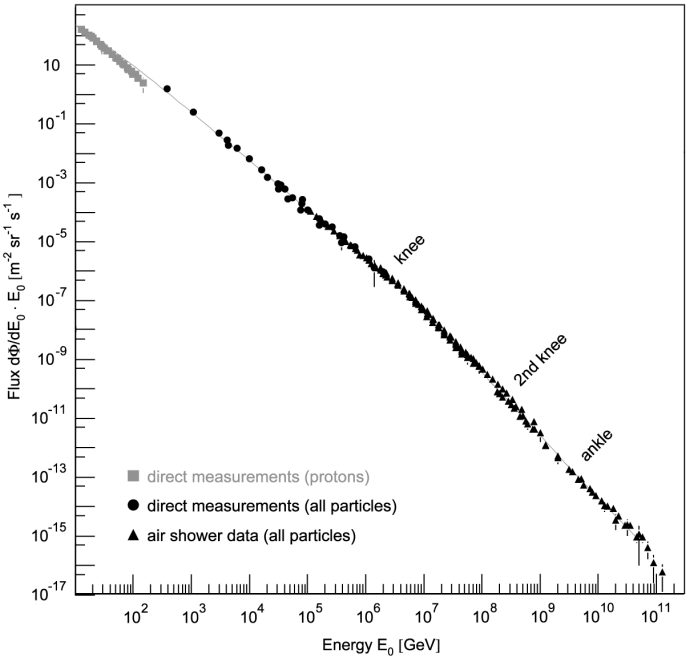


FIGURE 2.1: The Cosmic Ray energy spectrum, as measured by various detectors by direct measurement and indirectly from extensive air showers. The distribution follows a power law relation with a steepening beyond the 'knee'. Other structures in the spectra are labelled. Figure obtained from [1].

The Fermi Acceleration of cosmic rays produces a power law spectrum in energy, generally with a $\Gamma \approx 2$ spectral index. This in turn produces, via the decay and radiative processes described above, steeper power law spectrum of gamma rays. While absorption, cooling, and other factors affect the flux spectrum of a source, it is a well motivated assumption that they are typically power law relations.

2.2 Types of Sources

Following the discussion of the production of gamma rays, the various classes of gamma ray sources that utilise these processes will be introduced.

2.2.1 Active Galactic Nuclei

At the center of most galaxies, it is widely understood, lie a supermassive black hole with masses greater than $10^6 M_\odot$. Some of these black holes accrete material from surrounding gas and dust, and in extreme circumstances relativistic jets can form along the rotation axis of the accretion material[2]. The strong magnetic fields can accelerate electrons that produce synchrotron emissions. These photons, as well as photons that come from internal processes of the accretion disk[25], are then inverse Compton scattered to GeV energy ranges.

There have been many AGN classes discovered throughout the history of gamma ray astronomy. However they are understood to arise from a more general 'unified scheme' with both observational and intrinsic differences producing the various classes[26]. The scheme is generally broken down into radio loud and quiet schemes, characterising the strength of the spectra in the radio band. The various sub categories of these schemes depend on the orientation of the AGN to the observer. Blazars (radio loud) for example are observed when the AGN jets are close to parallel to the line of sight, while Seyfert I and II (radio quiet) galaxies are seen at various inclinations at which various emissions from the surrounding matter are seen. A general scheme can be seen below in the following figure.

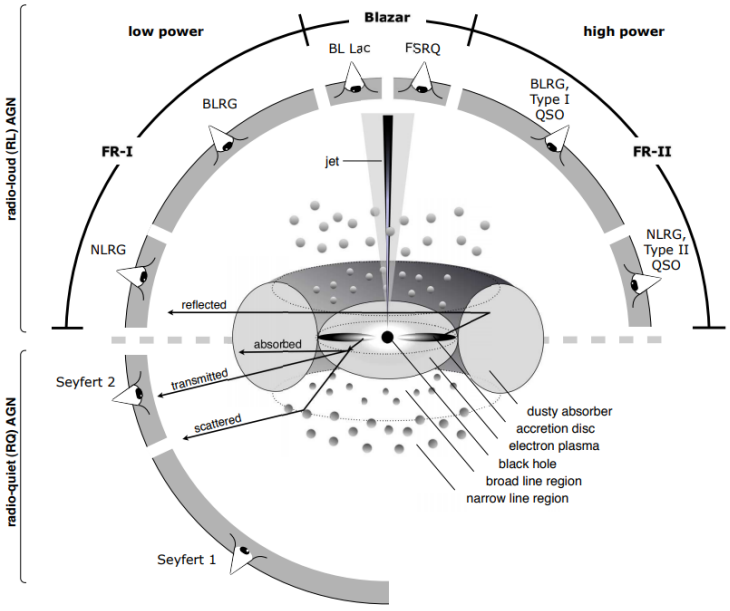


FIGURE 2.2: Diagram of the currently understood AGN unified scheme. The top half represent the Radio Loud AGN, the lower half representing the Radio Quiet AGN. The presence of jets, emission spectra, and inclination to the observer further categorise the scheme. Figure obtained from [2].

2.2.2 Pulsars

Pulsars are the largest population of galactic sources, and the second largest population of point sources in the Fermi catalogue[8]. This class of sources consists of neutron stars that rotate rapidly with extreme magnetic fields on the order of 10^8 Gauss and beyond. While radio astronomy has studied pulsars rotating with periods of a few seconds since the 1960's[27], more recently pulsars rotating with milli-second periods have been studied by the Fermi telescope[28].

The 'gap model' provides the general description of how pulsars produce highly accelerated charged particles and gamma rays[29]. The model relies on a relative vacuum existing in the magnetosphere, in which the electric and magnetic fields are perpendicular. The energy potential is able to accelerate the charged particles to high energies. At these relativistic speeds, the particles interact with the magnetic fields producing pair cascades and synchrotron radiation.

A number of versions of the gap model exist, differentiated by the location of the unshielded gap in the magnetosphere[30][31][32]. The inclination between the rotation axis and the magnetic moment axis creates a lighthouse effect, in which the misalignment creates pulses as the beam of gamma rays crosses the line of sight of the observer. This being the reason 'pulse' is the prefix of the name pulsar.

Above a few GeV's, the spectrum of a typical pulsar falls off with gamma ray emissions limited to low energy. However, the spectrum can be very luminous in the lower band. The lighthouse effect can be measured using temporal variability studies, but averaged over many rotation periods the observed flux is relatively steady.

2.2.3 Galactic Diffuse Background

The interstellar medium is a dominant source of gamma rays in the sky due to the interaction of cosmic rays with the gas/dust and radiation fields[33]. The spectrum of the diffuse gamma rays produced depends strongly on both the cosmic ray energy spectrum and the density of the various fields of the interstellar medium.

At the lower end of the HE gamma ray band, the primary interactions are electronic: bremsstrahlung in the gas fields, and the inverse Compton scattering of ambient light. Above this, inelastic scattering of HE protons produces all varieties of pions, of which the neutral pion primarily decays into two gamma rays. Inverse Compton scattering of the interstellar radiation field contributes to the very high energy part of the spectrum.

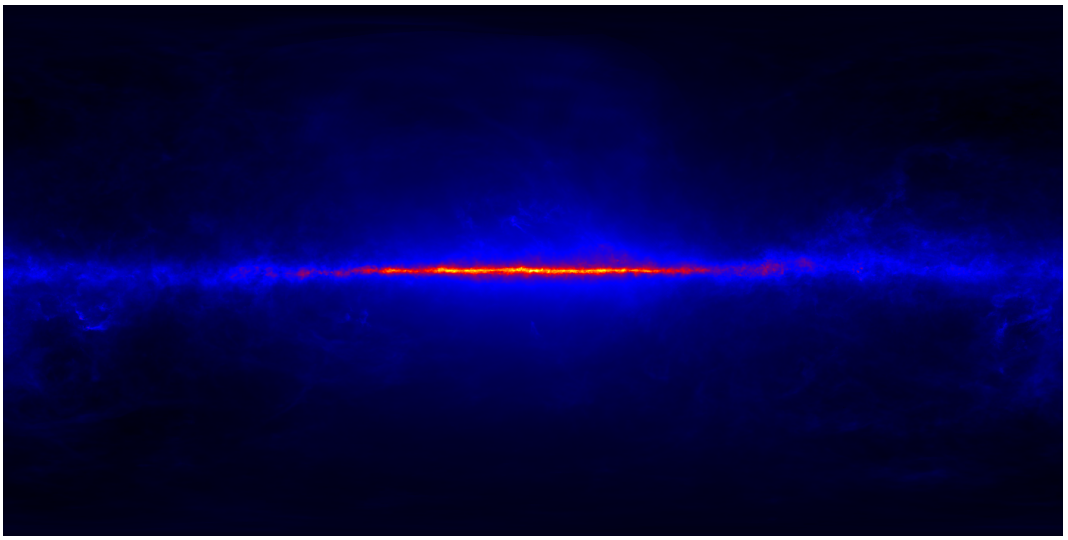


FIGURE 2.3: Galactic diffuse background as modelled for Pass 8 data. The model is for the energy bin centered at 58 MeV, and the coordinates follow a equirectangular projection.

In order to model the diffuse background, a knowledge of the density of the cosmic ray particle flux and the density of the interstellar gas/dust and radiation fields of the Milky Way is required. The gas fields can be studied and modelled as a linear combination of the density maps derived from CO and HI line surveys and dust maps[33]. On the other hand, cosmic ray injection and propagation are modelled using the GALPROP program[1] that computes the dispersion and energy gains and losses as they disperse around the galaxy. Together they allow a full model of the Galactic diffuse background to be formed, which is shown above.

2.2.4 Isotropic Diffuse Background

Diffuse gamma rays from extra-galactic sources, albeit weakly, can contribute to Fermi observations. Along with residual misclassified cosmic ray emissions, they make the Isotropic Diffuse Background. It has been suggested that unresolved blazar sources could be the major contributor to this background[24]. It is assumed to be isotropic with no significant variations, with a spectrum derived from an all sky emission fit to events not included in the Galactic Interstellar Model.

2.2.5 Earth Limb

The Earth's atmosphere acts as a gamma ray source due to the induced emissions from cosmic ray interactions[3]. In orbit, the Fermi telescope sees the Earth as the brightest

[1] galprop.stanford.edu/

source of gamma rays, purely because of how close it is. As it is an induced property, this is often referred to as the Earth albedo gamma ray emission.

Cosmic Rays collide with molecules of the atmosphere producing cascades of particles including both neutral and charged pions, kaons, and electrons. The decay of these particles, and bremsstrahlung of electrons, continue the cascade, while also emitting gamma rays. At the energies of the interactions that produce gamma rays, the cross sections are pronounced in the forward direction. These gamma rays can be detected by the Fermi LAT, though for most studies they are considered nuisance data.

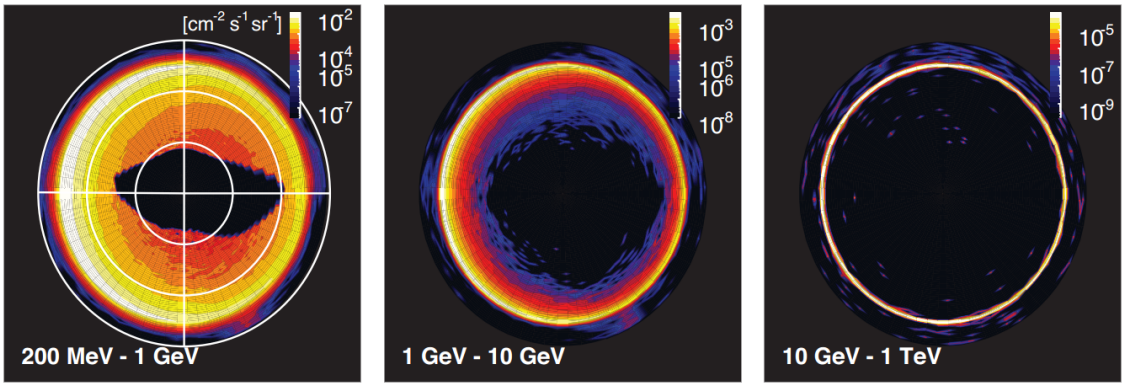


FIGURE 2.4: Intensity of the Earth Limb gamma ray spectrum in three energy bands. The diagrams are shown in polar coordinates, the center being the nadir direction and the edge being the horizon. Figure obtained from [3].

The inner part of the Earth's disk produces a soft gamma ray spectrum. Secondary particles of the cascade can be back scattered, and so their decay produces gamma rays away from the Earth (and the continuing cascade). However, on the edges of the Earths disk, the cosmic ray showers graze the relatively thin atmosphere. These gamma rays are more numerous and thus the limb of the atmosphere is a bright gamma ray source for Fermi. The limb region can be described by a power law spectrum of slope $\Gamma = 2.79 \pm 0.06$[3].

2.2.6 Sun

In the quiet periods of solar activity, the quiescent Sun is known to be a source of high energy gamma rays[34]. This is achieved in two ways. The outer layer solar photosphere interacts with cosmic rays to produce pion decays and subsequent gamma rays as has been described elsewhere. The second interaction is an Inverse Compton effect of photons in a spatially extended region around the Sun. Photons emitted by the Sun gain energy as they scatter from cosmic ray electrons.

Solar flares are often strong enough to produce both X-rays and gamma rays[35]. Magnetic fields in the Corona, through reconnection, release energy that accelerate electrons and plasma to relativistic energies. This plasma then interacts with surrounding protons producing pions. The neutral pions decay into gamma rays, while charged pions decay through a series of reactions to electrons and neutrinos which emit gamma rays via bremsstrahlung. This produces a continuous gamma ray spectra which lasts for a short period.

2.2.7 Moon

Much like the Earth's atmosphere, the Moon is a source of gamma ray emission. Cosmic rays in this case interact with the lunar surface, owing to the lack of an atmosphere and magnetic field of the moon[36]. There are two types of emissions, a continuous high energy spectra and nuclear emission lines.

The continuous spectra, first observed by EGRET, arises from cosmic rays that impact the surface and cascade into secondary particles. Emission occurs when the pion and kaon particles decay, or when electrons decelerate via bremsstrahlung. As with the Earth limb, the limb of the Moon is relatively brighter due to the cascades occurring tangentially. The surface material of the Moon (regolith) can emit nuclear emission lines in the gamma ray spectrum as well. Spallation and secondary particles from cascades can activate the surface material, and because of its composition, produce HE emission lines.

2.3 Gamma Ray Telescopes

The design of gamma ray telescopes is constrained by the interactions between gamma rays and matter. While visible light, used in optical astronomy, can be reflected and refracted onto a detector, gamma rays can not. Gamma rays interact strongly and penetrate deep into the mirror surface, and an image is unable to form. Detection of these high energy photons must thus be achieved in another manner. This section will motivate the design of these telescopes through the discussion of the gamma ray/matter interactions.

The first point to consider is that the atmosphere is opaque to the gamma ray band, and in fact is opaque above the ultra-violet frequencies. Unlike visible light and radio waves that are much less absorbed by the atmosphere and can be detected at ground level. This essentially forces direct gamma ray astronomy higher to balloon experiments or to

space observatories. Ground based Cherenkov detectors, such as the HESS project, are able to measure gamma ray signals at the TeV scale through the induced extensive air showers they produce in the atmosphere[37].

The interactions of light onto a metal is dependent on the energy of the incident light, with various processes contributing to the cross section of a specific metal. The photo-electric effect dominates the cross section at low energy where classic electrodynamics describes the interaction. The Compton scattering cross section peaks in the low energy band. While above 1 MeV, twice the rest mass of the electron, gamma rays can interact with the Coulomb field of the atomic nuclei and produce an electron-positron pair. Above 10 MeV pair production dominates all other processes.

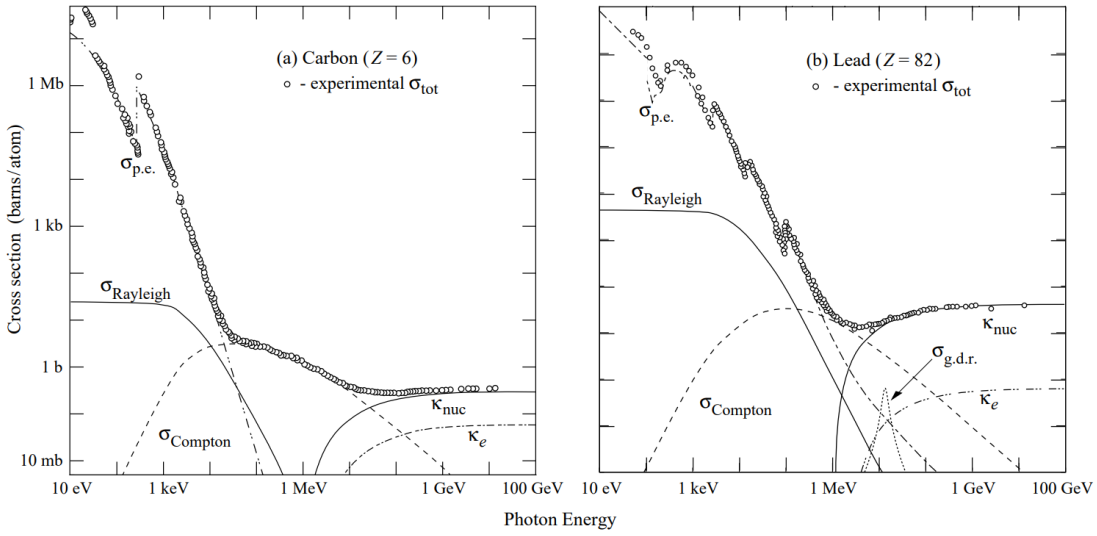


FIGURE 2.5: The various contributions to the cross section of a photon scattering from carbon and lead, in the HE gamma ray band. Experimental data, shown as circles, are compared to the theoretical distributions from the photoelectric effect, Rayleigh and Compton scattering, and pair production. Figure obtained from [4].

The above figure shows the cross section amplitude for both carbon and lead metal. Both show the same general trends, as described above, but some notable differences are present. In high-Z materials, such as lead ($Z = 82$), the cross section is an order of magnitude greater than carbon for high energy photons. The other difference of importance for gamma ray telescopes, is that the pair-production cross section rises faster for lead than carbon. Above 100 MeV, Compton scattering only accounts for a few percent of photon interactions.

The interactions in the HE band mean that for useful gamma ray flux measurements, a gamma ray telescope must essentially be a pair-production telescope. The electron-positron pair that are produced then interact themselves with the surrounding matter. The following will briefly introduce those interactions.

As has been mentioned earlier, the dominant interaction of electrons in matter is through bremsstrahlung radiation via the Coulomb fields of the nearby nuclei. The photon produced will then undergo pair-production itself. Both processes suffer an energy loss that is, to first order approximations, linear with distance[1].

$$\frac{dE}{dx} = \frac{E}{X_c} \tag{2.1}$$

X_c is a characteristic length scale associated with the energy loss. For bremsstrahlung radiation, it is taken to be the distance travelled in which the electron emits all but e^{-1} of its initial energy. This is referred to as the radiation length, X_0. For gamma rays in pair production, the mean free path between interactions is $\frac{7}{9}X_0$[4].

This cycle of electron producing bremsstrahlung radiation, in turn producing an electron-positron pair, in turn each producing bremsstrahlung radiation, etc, forms an electromagnetic shower. The shower grows exponentially with each pair production taking place, and will continue until the two processes reach their respective critical energy limits. For photons, this occurs when there is insufficient energy to produce an electron and positron (i.e. twice the rest mass m_e). Likewise, for electrons, the process can no longer be produced when the energy loss from ionisation is greater than the energy loss from bremsstrahlung. While this is material-dependent, for high-Z materials the limiting energy is about 10 MeV[4].

The electromagnetic shower develops in a tight cone around the intial gamma ray trajectory. Due to the electron-positron pairs being highly relativistic, the shower is strongly peaked at the center of the cone. However, perturbations from distant nuclei affect the path of the electrons. This multiple Coulomb scattering can be described as a stochastic random walk. For a sufficient thickness of material, the angle between the initial and perturbed trajectory follows a Gaussian profile with a standard deviation[4] following

$$\theta_0 = 8^\circ \sqrt{\frac{x}{X_0}} \frac{100MeV}{E} \ . \tag{2.2}$$

The above discussion motivates a number of key components of a gamma ray telescope. In order for a telescope to convert the majority of incident photons to electron-positron pairs, the telescope requires sufficient radiation lengths of material. Maximising the conversion process can be achieved with a high-Z material. Estimates of the direction and energy of the incident gamma ray must also be made. The direction can be estimated by measuring the first generation pair before multiple scattering takes place. While the energy can be estimated by converting and measuring the energy produced in the electromagnetic showers that develop in the converter material. The next section will describe how these principles were incorporated into the design of the Fermi LAT.

2.4 Fermi Telescope

The Fermi Gamma Ray Space Telescope, previously known as GLAST (Gamma Ray Large Area Space Telescope), was launched by NASA on the 11 June 2008 into a low-earth near-circular orbit of 565km and 26 degree inclination[38]. The project is an ongoing international collaboration between numerous space agencies, institutes and universities from France, Italy, Japan, Sweden, and the USA.

Built and designed on the principles of gamma ray astronomy discussed previously, the Large Area Telescope is able to perform all-sky surveys and be inertially pointed. The telescope is a valuable asset for research into high energy astrophysics.

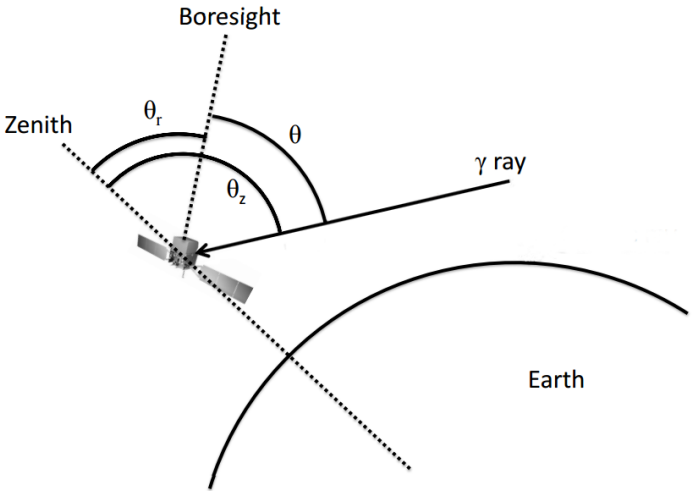


FIGURE 2.6: Detection of a gamma ray by the Fermi LAT. The rocking angle of the telescope is shown as θ_r, while the zenith and incidence angle of the gamma ray event are shown as θ_z and θ respectively. Figure adapted from [5].

Each orbit takes 3 hours to complete. During the default science mode, survey mode, the telescope orbits and rocks the boresight in alternate directions in alternate orbits. This achieves a near uniform exposure every two orbits. Initially this rocking angle was set to $\theta_r = 35°$, but now rocks at $\theta_r = 50°$.

In certain situations it may be decided (either by the telescope itself, or by ground control) to point towards a Target of Opportunity (TOO). This may be required when a study requires a higher exposure over a shorter time, such as pulsar timing studies or gamma ray burst studies. If the telescope (LAT or GBM) detect a gamma ray burst of significant brightness, then the telescope will automatically point towards the source.

2.4.1 Detector

The Fermi telescope was designed to be modular with 16 Tower Electronics Modules (TEM). Each TEM contains a Converter Tracker (abbreviated TKR), this is the primary module of layered material that converts gamma rays to electron-positron pairs. At the base of the module, a Calorimeter (CAL) measures the deposited energy from the electron-positron shower. An Anticoincidence Detector (ACD) protects the tracker system from charged particles giving false readings. The Data Acquisition System (DAQ) performs the readouts of each of these subsystems, controls the triggering responses, and data storage[38].

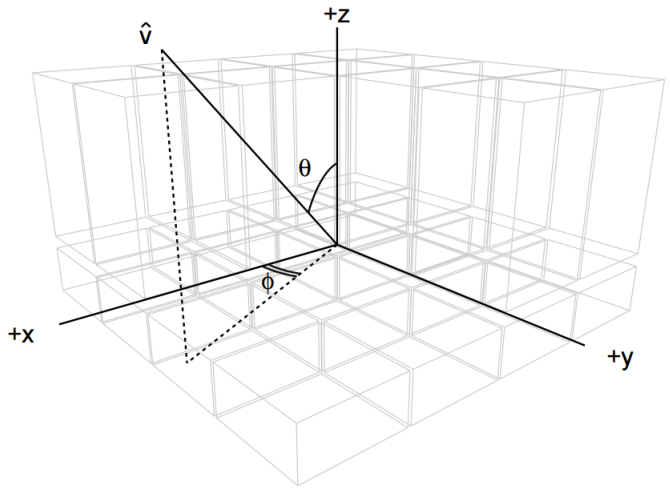


FIGURE 2.7: Diagram showing the layout of the Large Area Telescope. 16 calorimeter modules are shown with 12 (of the total 16) tracker modules on top. The ACD that covers the sides and top is not shown in this diagram. Figure obtained from [6].

Each TKR is made from tungsten, a high-Z converting material, and are woven with a series of sensitive detectors that can track the path of charged particles. The detectors are made from two layers of single sided silicon-strip detectors (SSD), one each to measure the 'x' and 'y' coordinates. Eighteen of these tracking planes are stacked together to form a TKR, and are held by trays made of carbon-composite material. Readout electronics are fitted to the side of the tray and connected to the DAQ.

The TKR's are divided into a Front region, the top twelve planes, and Back region, the next 4 planes. The bottom two planes do not contain a converting material. Front trackers have thin tungsten converters of 0.03 radiation lengths, in order to achieve a fine PSF at low energies, while the back trackers have converters that are 6 times thicker as a compromise between a less resolved PSF but a maximised effective area for higher energy photons.

In order to reconstruct the energy of an incident photon, the energy produced in the electromagnetic shower must be stored and measured. This is achieved with 16 calorimeter modules consisting of an arrangement of CsI(TI) optically isolated crystals. Together they provide a vertical depth of 8.6 radiation lengths, enough to measure energies up to the TeV scale.

The crystals scintillate due to the energy loss of the shower as ionisation to the CsI atoms. Photodiodes collect the light and allow the readout of the signal to the DAQ. The segmentation between crystals allows the electromagnetic shower profile to be imaged. This allows a reconstruction of the showers direction to be estimated and provide a veto to background events.

Charged particles that are detected by the TKR and calorimeter modules can be discriminated from true gamma ray signals due to the shower profiles that form. However, cosmic ray electrons form shower profiles very similar to the cascades that form from gamma rays and will not be rejected by the DAQ. The Anticoincidence Detector are a series of plastic scintillator tiles that surround the LAT and detect the charged particles that would otherwise go undetected. Averaged over the ACD, the charged particle detection has at least 99.97% efficiency rate.

2.4.2 Event Reconstruction

Data from the various subsystems of the LAT are considered 'Level 0' data when downlinked from the Fermi satellite. 'Level 1' data, or data usable by researchers, are the reconstructions of detected events in the LAT. Raw measurements and readouts from the tracker and calorimeter modules, and the anticoincidence detector, are processed together to create a unified event hypothesis. An event reconstruction results in around 1000 parameters as part of an hypothesis.

Reconstruction of a event's type and characteristics have been achieved in different ways throughout the life of the Fermi mission. The current reconstruction analysis is referred to as 'Pass 8'. The previous passes of reconstruction have utilised various Monte Carlo studies of the LAT instruments to improve the understanding of events and the Level 0 data they should produce. Pass 8 has continued this, while also providing new algorithms for track and energy recognition, and an overhaul of the event classification scheme.

2.4.3 Mission Goals

The Fermi telescope was designed with a mission lifetime of at least 5 years, with the goal of achieving 10 years. At the time of writing this thesis, the mission has run successfully for nine years and is continuing to collect data.

Being a successor to EGRET and other recent experiments, the Fermi telescope project includes goals that addressed the findings of EGRET that were to be studied more. Among other objectives, Fermi was also to develop an understanding of the physics of astrophysical sources and to probe new areas of physics, notably dark matter.

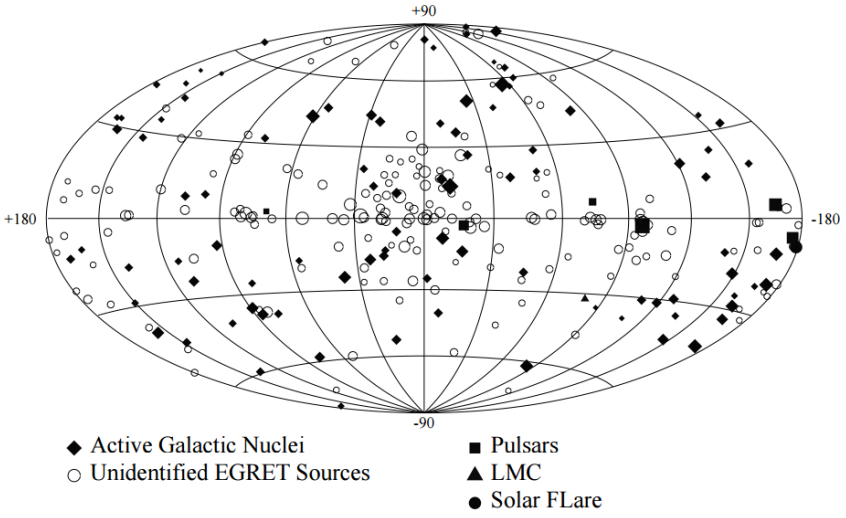
EGRET was able to identify 271 sources in the gamma ray sky, and of these 101 were categorised as pulsars and blazars due to the spectral properties derived from their analysis[7]. The remaining identified sources could not be categorised.

With the improvement in resolving power and source localisation, and being able to measure spectra across larger energy ranges, the Fermi telescope has been able to find many more sources and motivate population studies of source classes. Temporal changes in flux can also be monitored. The latest Fermi Source Catalogue (3FGL) contains 3033 sources, including detailed models of the diffuse galactic background[8]. A comparison between the EGRET catalogue and the 3FGL catalogue is shown in Figure 2.8.

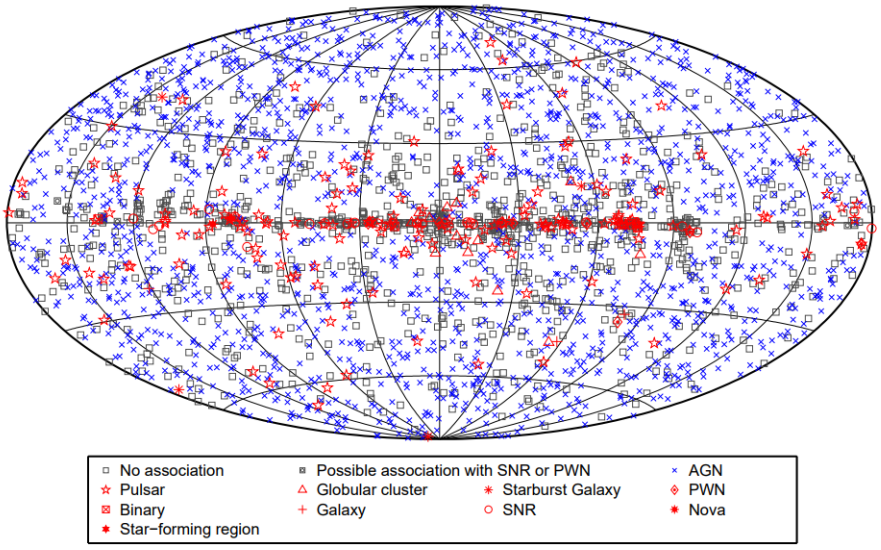
2.4.3.1 Mechanisms of particle acceleration

In previous sections, the way in which gamma rays are produced from relativistic particles has been described. Fermi Acceleration was presented as a likely explanation to the process that achieves this acceleration, however there are of course other non-thermal mechanisms that are less understood.

Variability in flux and the spectra from AGN sources can now be studied and may provide further evidence of the unification scheme and other models of AGN development, such as the blazar sequence that has been suggested. Relativistic jets produced in certain AGN convert the gravitational potential energy[39] of accreting matter, however the process is not well understood. Population studies of these jets could provide a more clear view of the formation and emission mechanisms involved[40]. Pulsar gap models are also candidates for detailed studies, in which the broad energy spectrum that can be studied may provide evidence more towards certain gap models than others[41].



(a) Third EGRET Catalog (1999)



(b) Third Fermi Catalog (2015)

FIGURE 2.8: Comparison of the gamma ray sources, shown in galactic coordinates, detected by both the EGRET[7] and Fermi[8] missions. The respective classification schemes are shown for each map.

2.4.3.2 Dark Matter Models

There is strong evidence that has emerged in recent decades of non-baryonic matter that composes a large amount of the total mass-energy in the universe, though its nature is largely unknown[42]. While a number of theories have been presented to describe dark matter and its interactions, one of the strongest candidates is a theory of WIMPs (Weakly Interacting Massive Particles). These types of particles interact only through the weak force and gravity, and typically have masses much larger than

those found in the standard model. Extensions of the standard model of particle physics through supersymmetry, extensions of the Higgs theory, and the inclusion of extra spatial dimensions, predict that WIMPs should have certain detectable properties[18].

It may also be possible to indirectly detect dark matter through the annihilation of WIMPs. This annihilation of two WIMPs can, through various decay and interaction pathways, lead to the production of gamma rays. Below the mass of the annihilating WIMP, a continuous spectrum could be produced from the decay of π^0 mesons. Also possible are a weaker signal of mono-energetic gamma rays produced when a two-particle final state is achieved (two photons, or a photon and a Z boson).

The Galactic Center is one the attractive regions to search for these signals, and it is believed that the LAT has the ability to detect them. However, many sources and uncertainty in the modelling of the diffuse background make it difficult to distinguish the annihilation signal.

Chapter 3

Model Derivation

While the various sources described in the previous Chapter all produce gamma rays by different mechanisms, they can each be modelled mathematically in a general framework describing the particle flux density. This description models both the spatial and spectral properties of individual sources. The functions will be parameterised in terms of direction and energy, neglecting the time dependence of variable sources for this study.

In typical analyses, physical quantities are often estimated directly from data. The visible band flux of a star, for example, can be estimated by aperture photometry. The resolving power in the visible band means that pixels illuminated by the stars image, only contain light from the star. However, this can not be achieved by the Fermi LAT. A large PSF at low energies, and the lack of isolated sources, means that the original source of a photon detection is often not clear.

The Instrument Response Functions of the LAT describe this uncertainty in the spatial origin and other event characteristics. The functions represent the probabilities of the dispersion of an observed value from its true value, and the efficiency at reconstructing gamma ray events. Extensive Monte Carlo simulations and on-board calculations have developed empirical relations for the IRFs. Pass 8, the latest version of these functions, will be used throughout this study (P8R2_v6).

By folding the particle flux density models through the response functions, and integrating over suitable ranges of interest, one can derive the expected number of counts, as a function of energy and direction, from the gamma ray model. This quantity can be directly compared to Fermi data and, though in a convoluted way, allow the estimation of model parameters.

3.1 Source Models

A gamma ray source is best described by the particle flux density. This value represents the rate at which photons from the source are incident per unit energy/area/time from the solid angle $d\vec{p}$ about the direction $\vec{p}$. This will be denoted by $S(E, \vec{p})$ throughout this study and typically be given the units photon m^{-2} MeV^{-1} s^{-1}. There is no time dependence of the particle flux density as it has been mentioned that any variable source will be steady over the time frames considered here.

Due to the angular resolution of the LAT, most gamma ray sources are identified as point sources. However, some sources are large enough or close enough to have an angular size larger than the resolving power. In the 3FGL source catalogue, 25 sources were modelled as extended sources in this way[8]. They were often identified when multiple clustered detections of 'sources' for the same source.

The particle flux density is often assumed to be factorised into two components, a spectrum dependent on energy and a spatial morphology dependent on position. The two factors being independent of each other.

$$S(E, \vec{p}) = \frac{dN(E)}{dE} \; f(\vec{p}) \tag{3.1}$$

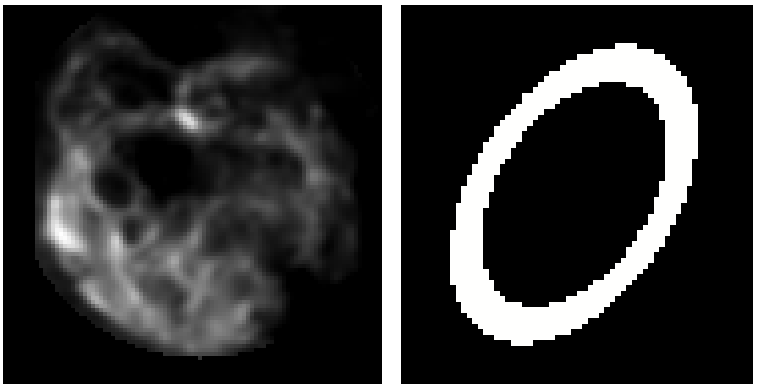
This factorisation is self-evident for point sources as they are treated as a localised source at specific coordinates with an emanating flux. The difference in particle flux across an extended source is typically only due to changes in intensity and not due to changes in the spectra. For more complex situations, such as the Galactic diffuse background, the assumption does not hold and instead the model is not factorised at all.

The intensity of the energy spectrum, or spatial morphology, is defined such that it is normalised when integrated over the entire sky

$$\iint_{\vec{p}} f(\vec{p}) d\vec{p} = 1 \; . \tag{3.2}$$

Point sources are thus expressed as a Dirac delta function $f(\vec{p}) = \delta(\vec{p} - \vec{p}_0)$ centered on the position $\vec{p}_0$. Extended sources, on the other hand, require more descriptive functions. In some instances these can be expressed by analytic descriptions such as disks, rings and 2-dimensional Gaussian surfaces. While others require specific maps derived in other ways. Shown in Figure 3.1, are two examples of the spatial functions used.

In many cases, the processes that produce gamma ray emissions form broadband spectra with power law type relations. These relations typically have a normalisation parameter and a number of shape parameters. While the Fermi source catalogues have introduced



(a) S 147 (b) W44

FIGURE 3.1: Extended Source templates from the 3FGL catalogue for S 147, expressed by a map, and W44, showing the ring structure. W44 has a constant intensity within the ring, while S 147 has a varying intensity depending on the position within the source.

shape parameters to account for steeper tails, more than modest curvature, and other structural differences, the simple power law model remains a useful descriptor. In individual small energy bands, it is often assumed that the spectrum will follow power laws inside the band.

Therefore, the power law will be defined as

$$\frac{dN(E)}{dE} = N_0 \left(\frac{E}{E_0} \right)^{-\Gamma} \quad , \tag{3.3}$$

where N_0 is the normalisation parameter, Γ is the spectral index, and E_0 is a scale parameter.

Alternatively, a power law may be defined by the flux F between E_{min} and E_{max},

$$\frac{dN(E)}{dE} = \frac{F(1-\Gamma)}{E_{max}^{1-\Gamma} - E_{min}^{1-\Gamma}} E^{-\Gamma} \ . \tag{3.4}$$

In general, the total model will be a sum of individual source models due to the additive nature of particle flux densities. In most cases this will include the Galactic diffuse background, the isotropic background, N_P point sources, and N_E extended sources.

$$S(E, \vec{p}) = S_G(E, \vec{p}) + S_I(E) + \sum_i^{N_P} \frac{dN_i(E)}{dE} \delta(\vec{p} - \vec{p}_i) + \sum_l^{N_E} \frac{dN_l(E)}{dE} f_l(\vec{p}) \tag{3.5}$$

3.2 Instrument Response Functions

When the various subsystems of the Fermi LAT are triggered by some detection, the DAQ will perform readouts of these subsystems and collect the data as a single event. Analysis and reconstruction of this event, accounting for the tracking with the TKR modules, the energy measurements of the shower profile, and the ACD trigger data, creates estimates of important parameters such as the direction and energy of the incident photon.

The way in which these reconstructions are performed have varied over the nine year history of Fermi. Currently, the Pass 8 software performs the reconstructions. As well as the standard direction and energy estimates, the reconstruction categorises events by an 'event class' and 'conversion region'. The event class groups events based on the quality of the reconstruction and the probability of being a photon and not a misclassified charged particle. For most analyses, including this study, the SOURCE class is used, providing a compromise between the quality of the parameter estimates and sensitivity of sources. The conversion region describes in which part of the TKR module the photon was converted to an electron-positron pair. These are either Front or Back regions, and was described in the previous Chapter. Pass 8 has also introduced alternatives to the Front/Back system, focusing on cuts based on the quality of direction or energy fits. However, they will not be used in this study.

Each pair of event class and conversion region have their own Instrument Response Functions (IRF). These functions describe the efficiency and quality of the reconstruction of an event as probabilities between the observed and true values of parameters. The IRF is denoted by $R(E', \vec{p}'; E, \vec{p})$, a function of observed variables (primed) given the true values of those variables (unprimed).

It is assumed that the IRF can be factorised into three individual functions that explain the instruments response to different parameter reconstructions. These are the effective area $A_{\text{eff}}(E, \vec{p})$, the point spread function (PSF) $P(\vec{p}'; E, \vec{p})$, and the energy dispersion $D(E'; E, \vec{p})$.

$$R(E', \vec{p}'; E, \vec{p}) = A_{\text{eff}}(E, \vec{p})\ P(\vec{p}'; E, \vec{p})\ D(E'; E, \vec{p}) \tag{3.6}$$

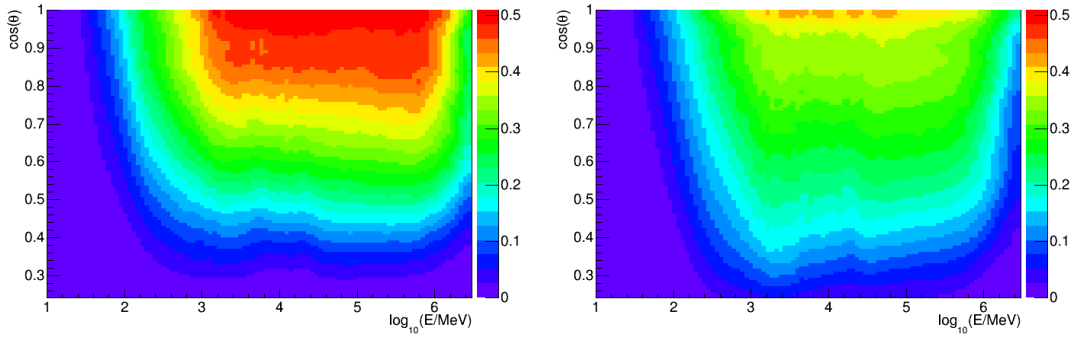
Monte Carlo simulations of the reconstructions of events based on simulated LAT Level 0 data, provides the empirical functions that make the IRF. The PSF and effective area results agree with data collected by Fermi in orbit, and are corrected based on discrepancies in some instances. The energy dispersion can not be verified by in-orbit measurements as there are no monoenergetic sources to calibrate from.

3.2.1 Effective Area

The effective area represents the collecting area of the telescope for an inclined photon and the ability to detect a gamma ray. The latter term describes the probability that the incident gamma ray begins a pair-production cascade in the TKR module, is readout by the DAQ, and that the event passes the event class selection cuts.

Factors such as the deadtime from the readout process, ghost tracks from charged particles just before the readout, and the reconstruction efficiency all affect the effective area. This creates a complex set of relationships and dependencies that cannot be modelled as analytic functions.

Instead, the effective area remains as an empirical table with a value quoted for each pair of energy and direction. While the dependence on direction has been quoted as $\vec{p}$, in most cases only the angle from the boresight, θ, is considered. Any dependence from the azimuthal angle, ϕ, is averaged out over long time frames.



(a) Front Conversion (b) Back Conversion

FIGURE 3.2: Colour plots showing the effective area for Front and Back conversions and the dependence on the energy and inclination of the incident photon. The colour axis is measured in m^2. Figure obtained from [9].

The Field of View (FOV) is commonly quoted as a key characteristic of the performance of the Fermi LAT and is defined as,

$$FoV(E) = \frac{\int A_{eff}(E,\theta)d\Omega}{A_{eff}(E,\theta=0)} \tag{3.7}$$

At 1 GeV, the Field of View is quoted as 2.4 steradian and is considered to be a wide view.

3.2.2 Point Spread Function

The PSF arises due to the uncertainty in reconstructing a photon event. It is expressed as the probability of measuring a photon in the solid angle $d\vec{p}$ at position $\vec{p}'$ given that it originated from $\vec{p}$ with energy E.

When a gamma ray penetrates the tracker and is converted to an electron-positron pair, the pair may undergo multiple scattering within the detector. This process obscures the reconstruction process and makes it difficult to provide useful estimations. This is particularly prevalent at low energies ($\sim$ 100 MeV) and is expected to drop at a rate of E^{-1}. At higher energies, multiple scatterings are less problematic and instead event reconstruction is limited by the resolution of the SSD strips[6].

The angular deviation between observed and reconstructed directions is defined as

$$\delta\vec{p} = 2\arcsin\left(\frac{|\vec{p}' - \vec{p}|}{2}\right) \tag{3.8}$$

and is further scaled by an energy dependent term $S_P(E)$, such that $x = \frac{\delta\vec{p}}{S_P(E)}$. This term is given by[10],

$$S_P(E) = \sqrt{\left[c_o\left(\frac{E}{100MeV}\right)^{-\beta}\right]^2 + c_1^2}\ . \tag{3.9}$$

This scale factor accounts for the first-order energy dependence of the PSF. While c_0 and c_1 vary depending on the event type used, the β term is held fixed at 0.8.

Results from extensive simulations of spatial reconstruction are binned in both energy and inclination angle. An example of the scaled angular deviation plotted as a histogram is shown below in Figure 3.3 for the bin centered on 7.5 GeV and 30 degree inclination of front conversions.

The resulting distribution can be parametrised by a sum of two King Functions. These functions, originally developed to describe many-body gravitational dynamics[43], are a generalisation of a Gaussian curve. The King function has been used previously to describe the PSF of the XMM-Newton mission launched in 1999[44].

The un-normalised King Function $K(x)$, is defined as such,

$$K(x) = \left(1 + \frac{u}{\gamma}\right)^{-\gamma}\ . \tag{3.10}$$

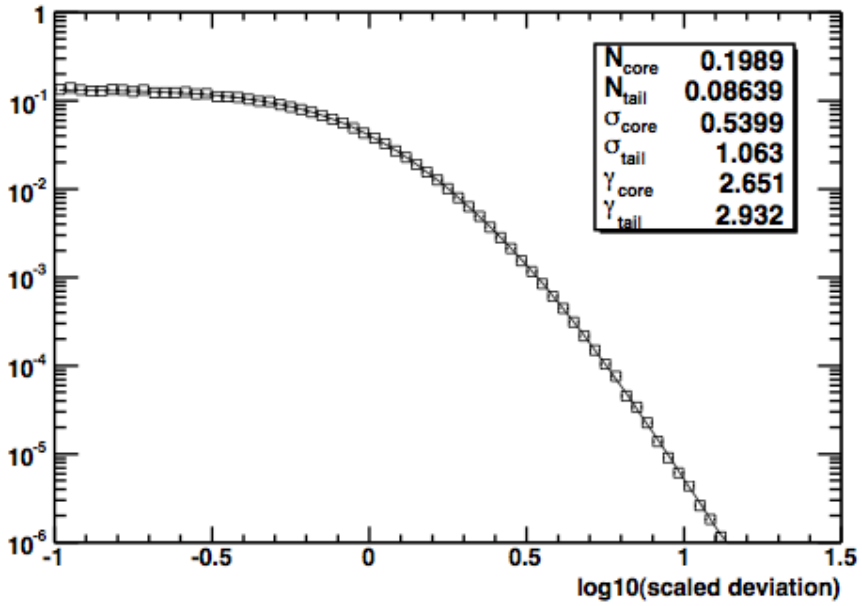


FIGURE 3.3: Density histogram of the scaled angular deviation of the reconstruction of Monte Carlo events. These simulations are drawn from front converted events in a bin centered on 7.5 GeV and 30 degree inclination. The distribution has been fit with two King functions, whose parameters are shown in the right-hand corner. The vertical axis is an arbitrary scale. Figure obtained from [10].

In taking the limit as $\gamma \to \infty$ and making the choice of $u \to u^2$, one returns the standard Gaussian function, $\exp(-u^2)$. Comparing the two distributions, the Gaussian tails fall off relatively faster whereas King function tails generally follow power laws of slope $-\gamma$.

The analytic description of the PSF of the Fermi LAT takes the form as such,

$$K(x, \sigma, \gamma) = \frac{1}{2\pi\sigma^2}\left(1 - \frac{1}{\gamma}\right)\left[1 + \frac{1}{2\gamma}\frac{x^2}{\sigma^2}\right]^{-\gamma} . \tag{3.11}$$

The King function is normalised over both the solid angle and the scaled deviation, $\int_0^\infty 2\pi x K(x, \sigma, \gamma) dx = 1$.

$$P(x) = f_{core} K(x, \sigma_{core}, \gamma_{core}) + (1 - f_{core}) K(x, \sigma_{tail}, \gamma_{tail}) \tag{3.12}$$

$$f_{core} = \frac{1}{1 + N_{tail}\left(\sigma_{tail}/\sigma_{core}\right)^2} \tag{3.13}$$

The IRF requires two King functions to properly explain both the tail and the core of the distribution. f_{core} represents the ratio between the core King function and the tail King function, while the σ_{core}, γ_{core}, σ_{tail}, and γ_{tail} parameters describe the respective

slopes and widths of the King functions themselves. A sixth arbitrary normalisation, N_{core} is also fitted.

3.2.3 Energy Dispersion

In a similar manner to the PSF, an events reconstructed energy estimate is subject to uncertainty. The dispersion function $D(E';E,\vec{p})$ is defined to be the probability of measuring a gamma ray with energy E' given that its true energy was E with a direction $\vec{p}$.

The deviation from the true value is given as $\delta E = E' - E$. This deviation is scaled by a function parabolic in $\log(E)$ and in $cos(\theta)$, with parameters fit from the Monte Carlo simulations that have been introduced earlier[45]. For brevity, these distributions will not be described here.

Across the energy range that Fermi has the capability of detecting, the energy dispersion only becomes significant in the low regime below 100 MeV. Above this, and not again until 1000 GeV, the bias lies mostly within 5% of an unbiased estimator. While detailed studies may require a full analysis including accounting for the energy dispersion, for most studies the cost of the additional computational effort is not worth the minimal bias that results from neglecting it. Throughout this study, energies considered will always be in the range (100 MeV to 10^6 MeV) that ensures minimal bias.

3.2.4 Time Dispersion

One can also consider the time dispersion in the IRF. However, as the LAT uses GPS systems to time measurements, the accuracy of measurements in time are on the order of micro seconds[38]. As only steady state models will be considered, and the accuracy is so high, time dispersion will be treated as a Dirac delta Function ($\delta(t'-t)$) and will be implied but not mentioned in the descriptions that follow.

3.3 Counts

The model that has been considered in previous sections has been developed around the particle flux density of various gamma ray sources. However, when the Fermi telescope takes measurements the results are recorded in terms of counts (or the particle flux). One must therefore fold the particle flux density through the IRF, before integrating

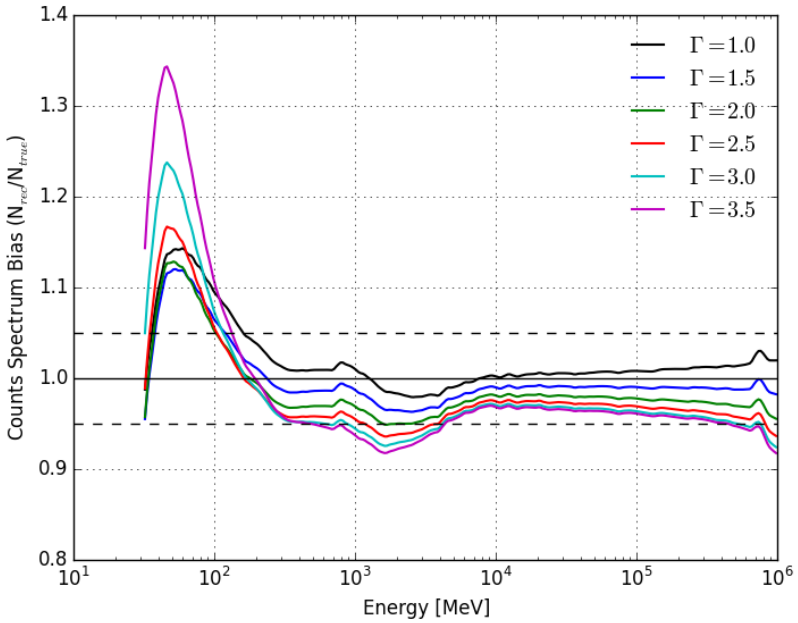


FIGURE 3.4: Bias in the flux of power law sources, of different spectral indices, when neglecting energy dispersion. The unbiased horizontal line at 1.0 is surrounded by 5% dashed margins. Figure obtained from [11].

over the true (unprimed) parameter values to arrive at the particle flux, in order to compare data to model.

The best way to form a model for some data set is to model the entire gamma ray sky. It should go without saying that this is wildly impractical and unfeasible for many studies. Contributions from sources on other sides of the celestial sphere are very small and negligible, and so it is convenient to define the Source Region (SR) and Region of Interest (ROI). This Source Region is the region of the sky in which sources are realistically able to contribute to the data in the Region of Interest. While not a fixed rule, the Source Region is recommended to extend at least 5° from the ROI for high energy sources, and 10° for sources dominated by low energy emissions.

Let $M(E', \vec{p}')$ be the counts rate in an infinitesimal bin within the ROI, and $S(E, \vec{p})$ be the particle flux density model for sources within the SR. Then,

$$M(E', \vec{p}') = \int_{SR} dE \; d\vec{p} \; R(E', \vec{p}'; E, \vec{p}) \; S(E, \vec{p}) \; . \tag{3.14}$$

The integral is performed over the energy range of the LAT and over the solid angle of the Source Region.

Fermi studies are binned spatially, and in many cases are binned in energy too. Unbinned energy analyses require the computation of further quantities and can drastically

affect the computational strain during modelling, especially for large numbers of events. This study will only consider data binned in energy and space. From the counts rate of an infinitesimal bin, one can compute the number of predicted counts in a bin by integrating over the measured energy and position of the respective bin, and further over the duration of the measurements.

Thus, the predicted number of photons measured in a finite sized bin (in energy and position) can be expressed as

$$\nu_i = \int_{bin_i} dE' \; d\vec{p}' \; dt \; M(E', \vec{p}') \; . \tag{3.15}$$

Substitution of the full expression of the infinitesimal counts rate, $M(E', \vec{p}')$ allows the exposure to be defined.

$$\begin{aligned} \nu_i &= \int_{bin_i} dE' \; d\vec{p}' \; dt \; \left(\int_{SR} dE \; d\vec{p} \; R(E', \vec{p}'; E, \vec{p}) \; S(E, \vec{p}) \right) \\ &= \int_{SR} dE \; d\vec{p} \; \left(\int_{bin_i} dE' \; d\vec{p}' \; dt \;\; R(E', \vec{p}'; E, \vec{p}) \right) \; S(E, \vec{p}) \end{aligned} \tag{3.16}$$

This model-independent term bracketed can be computed and expressed as the exposure, a function in terms of the true values of parameters. The exposure represents the product of area and time for a given position producing counts in the ROI. $\epsilon(E, \vec{p})$, is defined as

$$\epsilon(E, \vec{p}) = \int_{ROI} dE' \; d\vec{p}' \; dt \; R(E', \vec{p}'; E, \vec{p}) \; . \tag{3.17}$$

This finally derives a useful expression of the expected number of counts in a finite bin from a gamma ray model, that can be compared directly to simulated or collected data. This prompts the following Chapter on the best way to find the fitted parameters of the model that match the expected number of counts to the actual number from data.

$$\nu_i = \int_{SR} dE \; d\vec{p} \; \epsilon(E, \vec{p}) \; S(E, \vec{p}) \; . \tag{3.18}$$

Chapter 4

Maximum Likelihood Analysis

In the previous Chapter, a method of modelling gamma ray sources was motivated that takes into account both spectral shapes and spatial morphology. The model was then folded through the IRFs of the Fermi LAT, to arrive at the predicted number of photons in a finite bin (of energy and position). The model was described in terms of a number of parameters that characterise various gamma ray sources, but it was left vague how these parameters should be chosen when fitting a model to some data set.

The goal in finding these parameters is to infer the best choices of parameters that match the model to data. In previous photon counting experiments, statistics such as χ^2 fitting have been used to derive the best estimates of these parameters[46]. However, this statistic (and others) rely on the assumption that the number of counts follows a Gaussian distribution. In cases of low counts, this assumption is not necessarily achieved.

A more general approach has been developed [47][12] and used by a number of recent X-ray and gamma ray telescopes, including the Fermi telescope. The maximum likelihood method provides a convenient way of inferring model parameters, with little requirements on the underlying distributions. Subject to a number of regularity conditions, the maximum likelihood estimator (MLE) is an efficient and consistent estimator that is asymptotically normally distributed.

The likelihood function, that will be defined in this Chapter, describes the probability that the observed data could have originated from some underlying distribution for a given set of model parameters. By maximising this function with respect to the parameter space, one finds model parameter choices that yield the most likely estimates. For data considered in this study, it is assumed that within each bin that the source model produces photons according to a Poissonian distribution. The likelihood function will be derived for this case.

4.1 Principles of Maximum Likelihood

Consider an experiment that produces a data set, $\mathbf{x}$, that has arisen from an underlying probability density function, defined by parameters $\boldsymbol{\theta} = \{\theta_1, \theta_2, ..., \theta_k\}$, such that $X \sim f(\mathbf{x}|\boldsymbol{\theta})$ is the probability density function. The data is taken to be a vector, $\mathbf{x} = \{x_1, x_2, ..., x_n\}$, with n bins. The set of parameters $\boldsymbol{\theta}$ belong to the k-dimensional parameter space Ω.

In this way, $f(\mathbf{x}|\boldsymbol{\theta})$ represents the model or data generating process under consideration. It will be seen in later sections that this is implicitly related to the mathematical prescription that was derived in the previous Chapter.

As mentioned in the introduction, statistical inference is the estimation of the probability that the parameters of the distribution $f(\mathbf{x}|\boldsymbol{\theta})$ are conditional on the observed data set. This probability, $\xi(\boldsymbol{\theta}|\mathbf{x})$, can be related to the data generating process by Bayes theorem, which states

$$\xi(\boldsymbol{\theta}|\mathbf{x}) = \frac{f(\mathbf{x}|\boldsymbol{\theta})\xi(\boldsymbol{\theta})}{\int_\Omega f(\mathbf{x}|\boldsymbol{\theta})\xi(\boldsymbol{\theta})d\boldsymbol{\theta}} \ . \tag{4.1}$$

The posterior probability $\xi(\boldsymbol{\theta}|\mathbf{x})$ is thus proportional to the product of $f(\mathbf{x}|\boldsymbol{\theta})$ and the prior probability $\xi(\boldsymbol{\theta})$, as the denominator is simply a function of $\mathbf{x}$. In this form, the probabilities emphasize the difference between the Bayesian and Frequentist schools of inference. The maximum likelihood method is a frequentist method of inference. The parameters that form the model are considered as constant values in Frequentist inference, while in Bayesian inference they are considered random variables.

Equation 4.1 can be rewritten to include the likelihood function as $\xi(\boldsymbol{\theta}|\mathbf{x}) \propto L(\boldsymbol{\theta}|\mathbf{x})\xi(\boldsymbol{\theta})$, in which all the inference comes from the likelihood function.

The likelihood function, $L(\boldsymbol{\theta}|\mathbf{x})$, is defined to assign a value to each set of parameters $\boldsymbol{\theta}$ in the parameter space Ω so as to indicate how likely the parameters were to have generated the data $\mathbf{x}$. The likelihood is proportional to the joint probability of the data given the parameters $\boldsymbol{\theta}$. For simplicity, as the proportionality factor is dependent only on the data and can not be known easily, the proportionality is taken as equality. The data, $\mathbf{x}$, is a vector of n iid random variables, and so the joint density function can be written as the product of individual density functions for each variable.

$$L(\boldsymbol{\theta}|\mathbf{x}) = f(\mathbf{x}|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}) \tag{4.2}$$

As the likelihood is proportional to the joint probability function by a data dependent term, the likelihood function is not a probability distribution. It instead portrays the

relative uncertainty in the choice of parameters from Ω. Because the constant of proportionality is data dependent, the likelihood can only be compared within a data set and not between two different data sets.

Computationally, evaluating the products of small numbers is prone to round-off errors and other such problems. In preparation for both the practical applications and the statistical distributions of Chapter Six, it would be more convenient to work with the log-likelihood. As the log function is a monotonic transformation, the log-likelihood carries the same information as the likelihood function. This is defined as such,

$$l(\boldsymbol{\theta}|\mathbf{x}) = \log(L(\boldsymbol{\theta}|\mathbf{x})) = \sum_{i=1}^{n} \log(f(x_i|\boldsymbol{\theta})) \ . \tag{4.3}$$

Thus, the central principle of a maximum likelihood analysis is that the model parameter choices, $\hat{\boldsymbol{\theta}}_{\mathrm{MLE}} \in \Omega$, that maximise the likelihood (or equivalently, the log-likelihood) function are the best estimates. This can be expressed as

$$\hat{\boldsymbol{\theta}}_{\mathrm{MLE}} = \max_{\boldsymbol{\theta}\in\Omega} L(\boldsymbol{\theta}|\mathbf{x}) = \max_{\boldsymbol{\theta}\in\Omega} l(\boldsymbol{\theta}|\mathbf{x}) \ . \tag{4.4}$$

Put another way, the MLE parameters are a solution to the differential equation,

$$0 = \vec{\nabla} l(\boldsymbol{\theta}|\mathbf{x})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\mathrm{MLE}}} \tag{4.5}$$

Other solutions to the equation may exist that do not give the desired maximum likelihood. There may be local maxima, that while maximising the likelihood in a small region of parameter space, are not the global maximum in the entire parameter space. Saddle points, and local and global minima are also possible. However, the Hessian matrix can be used to check the solutions and evaluate if the global maximum has been found.

4.2 Regularity conditions

In order to ensure that the desirable properties of the maximum likelihood estimates are retained, a number of conditions and restrictions on the parameter space and the likelihood function must be made.

The sample space, S, defines the possible outcomes of the random variable, $\mathbf{x} \in S$, from the joint probability distribution $f(\mathbf{x}|\boldsymbol{\theta})$. The parameters of the distribution, $\boldsymbol{\theta}$, are members of a bound and closed set, the compact subset Ω. This will be referred to as the parameter space. For every unique choice of parameter, the corresponding

probability density function is positive definite ($f(\mathbf{x}|\boldsymbol{\theta}) > 0$ for all $\mathbf{x} \in S$) and also unique. That is, the choice of parameter determines the density function uniquely. Each probability density function must also have common support for all $\boldsymbol{\theta} \in \Omega$.

Let $\hat{\boldsymbol{\theta}}_{\text{MLE}}$ be the maximum likelihood estimate and $\boldsymbol{\theta}_0$ be the true parameters of the underlying model, such that $\hat{\boldsymbol{\theta}}_{\text{MLE}}, \boldsymbol{\theta}_0 \in \Omega$. The true parameters must be interior points of Ω, i.e. there exists a region around $\boldsymbol{\theta}_0$ that is completely within Ω.

These conditions guarantee that in the limit of large sample sizes, the maximum of the log-likelihood function is achieved by the true parameters. That is,

$$\lim_{n\to\infty} P_{\boldsymbol{\theta}_0} [\ l(\boldsymbol{\theta}_0|\mathbf{x}) > l(\boldsymbol{\theta}|\mathbf{x})\ ] = 1, \quad \text{for all } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \tag{4.6}$$

Conditions are also placed on the likelihood function in regards to differentiability. $l(\boldsymbol{\theta}|\mathbf{x})$ must be continuous in $\boldsymbol{\theta}$, and be twice differentiable in the neighborhood of $\boldsymbol{\theta}_0$. The third condition is that the order of integration and differntitiation is interchangeable with respect to $\boldsymbol{\theta}$.

Together, these restraints define the Fisher information and provide a relation to the variance of the score function. It can be shown that the score function, the derivative of the log-likelihood function with respect to $\boldsymbol{\theta}$, satisfies

$$\mathrm{E}\left[\frac{\partial\ l(\boldsymbol{\theta}|\mathbf{x})}{\partial\boldsymbol{\theta}}\right] = 0\ . \tag{4.7}$$

The Fisher Information $I(\boldsymbol{\theta})$ follows

$$\begin{aligned} I(\boldsymbol{\theta}) &= -\mathrm{E}\left[\frac{\partial^2\ l(\boldsymbol{\theta}|\mathbf{x})}{\partial^2\boldsymbol{\theta}}\right] \\ &= \mathrm{VAR}\left(\frac{\partial\ l(\boldsymbol{\theta}|\mathbf{x})}{\partial\boldsymbol{\theta}}\right)\ . \end{aligned} \tag{4.8}$$

An additional regularity condition is often required to ensure that an $\mathcal{O}(\boldsymbol{\theta}^3)$ Taylor expansion of the log-likelihood is bounded. This requires that the probability density function be three times differentiable and that there exists a function $M(x)$ such that

$$M(x) \geq \left|\frac{\partial^3\ l(\boldsymbol{\theta}|\mathbf{x})}{\partial^3\boldsymbol{\theta}}\right| \tag{4.9}$$

and a constant c that for each $\theta_0 - c < \theta < \theta_0 + c$ and all $\mathbf{x} \in S$, $\mathrm{E}_{\boldsymbol{\theta}_0}[M(X)] < \infty$.

4.3 Properties of the MLE

Due to the regularity conditions imposed on the distributions and spaces of the maximum likelihood method, the MLE can be shown to possess a number of important qualities that will be discussed in this section. However, the proof and full derivation of these properties will be omitted in this thesis as they are provided elsewhere.

In what is possibly the most important property, the MLE is a consistent estimator. As the sample size of data increases, the parameter estimates converge in probability[1] to the true parameters.

$$\hat{\boldsymbol{\theta}}_{\mathrm{MLE}} \to \boldsymbol{\theta}_0 \tag{4.10}$$

This is derived by demonstrating that the probability of a subset of S, for which the likelihood at the true parameters is greater than the likelihood of parameters in the neighborhood of $\boldsymbol{\theta}_0$, is bounded above and below by unity in the limit as $n \to \infty$. This, along with the first set of regularity conditions, are enough to prove a consistent estimator.

The further regularity conditions are used to show that the MLE is asymptotically normally distributed (converging in distribution[2]) about the true parameter $\boldsymbol{\theta}_0$. The variance of this distribution is given by the inverse of the Fisher information.

$$\hat{\boldsymbol{\theta}}_{\mathrm{MLE}} \sim \mathcal{N}\left(\boldsymbol{\theta}_0, I(\boldsymbol{\theta}_0)^{-1}\right) \tag{4.11}$$

The Fisher information is evaluated at the true parameter value, θ_0, in the above description. The consistency property ensures, however, that the MLE can be used in its place as in the limit of increasing sample size the MLE will approach $\boldsymbol{\theta}_0$.

By taking a Taylor expansion of the log-likelihoods first derivative about $\boldsymbol{\theta}_0$, the regularity conditions ensure that the $\mathcal{O}(\boldsymbol{\theta}^3)$ terms are bounded. As $\frac{\partial\, l(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}}$ is taken as a random variable, the Central Limit Theorem implies an asymptotic normal distribution. With suitable rearrangement of terms, one can arrive at the expression given above.

An extension of this property is that the MLE is an asymptotically efficient estimator. Efficiency of an estimator is a measure of the relative variance of the asymptotic normal distribution to the inverse of the Fisher information. In this case, the two are equal in the large sample size limit, giving an efficiency of $e\left(\hat{\boldsymbol{\theta}}_{\mathrm{MLE}}\right) \to 1$.

[1] A sequence of random variables X_n converges in probability to X^*, if for every $\epsilon > 0$ the expression $\lim_{n\to\infty} P(|X_n - X^*| < \epsilon) = 1$ is satisfied.

[2] A sequence of random variables X_n converges in distribution to X^*, if the Cumulative Distribution Function F_n converges to the Cumulative Distribution of X^* $\lim_{n\to\infty} F_n(x) = F^*(x)$.

4.4 Maximum Likelihood with Fermi

In the previous Chapter, the number of counts expected in some finite bin, labelled i, of energy and space was determined to be

$$\nu_i = \int_{SR} dE\ d\vec{p}\ \epsilon(E,\vec{p})\ S(E,\vec{p})\ . \tag{4.12}$$

The source model of particle flux density $S(E,\vec{p})$ was folded through the Instrument Response Functions, or equivalently the exposure, and integrated over suitable parameter ranges for the particular binning.

While individual measurements will vary due to sources emitting gamma rays independently of the time since the last gamma ray emission, the i^{th} pixel should detect ν_i counts during the time interval of interest on average. It is natural to assume that the probability of a number of counts occurring in this time interval will follow a Poisson distribution.

$$P(k_i \text{ counts}) = \frac{e^{-\nu_i}\nu_i^{k_i}}{k_i!} \tag{4.13}$$

For reasons of brevity, the dependence of ν_i on the parameters of the model are implied but not included.

The joint probability distribution that generates the sample space S, is thus a product of Poisson probability mass functions for each pixel. The log-likelihood for this situation can be derived as follows,

$$\begin{aligned} l(\boldsymbol{\theta}|\mathbf{x}) &= \sum_{i=1}^{n} \log(f(x_i|\theta)) = \sum_{i=1}^{n} \log\left(\frac{e^{-\nu_i}\nu_i^{k_i}}{k_i!}\right) \\ &= \sum_{i=1}^{n} \left(k_i \log(\nu_i) - \log(k_i!) - \nu_i\right) \\ &= \sum_{i=1}^{n} \left(k_i \log(\nu_i) - \log(k_i!)\right) - N \end{aligned} \tag{4.14}$$

where N is the total sum of the expected counts for all bins, $N = \sum_{i=1}^{n} \nu_i$. While not addressed in this study, if one takes the limit of infinitesimal bins that either hold 0 or 1 photon, the unbinned analysis is recovered.

The $\log(k_i!)$ term is only dependent on the data, not on the particular model fit by the MLE. When comparing the likelihood function within a data set for different models or choices of parameters, the term will always be the same. So in some instances this term is neglected from the calculations and the likelihood value quoted is of the form $l(\boldsymbol{\theta}|\mathbf{x}) = \sum_{i=1}^{n} k_i \log(\nu_i) - N$.

Chapter 5

FermiTools Analysis

A suite of tools were developed by the Fermi collaboration to provide specific commonly used functions in the analysis of data collected by the Fermi telescope. The tools are an extension of the FTOOLS environment. The FTOOLS package was designed for analysing FITS format data from a number of previous high energy astrophysics projects, but is now maintained by NASA HEARSARC (High Energy Astrophysics Science Archive Research Center). The Fermi Science Tools are publicly available and provided by the FSSC[1].

While individual studies of Fermi data may vary in scope and subject, each will use a number of common standard functions. As has been developed in the previous Chapter, the maximum likelihood analysis requires calculation of the exposure, and to bin raw data into pixels. These two processes can be achieved using FermiTool functions.

The tools that are included allow a researcher to preform a number of typical analyses while still allowing sufficient freedom in how these analyses are applied. The tools allow one to simulate the observations and orbit of the LAT, to select and bin data, and also to perform model fitting to data. They are often named with the prefix 'gt', such as '`gtlike`' for performing the likelihood fitting process. A full list of the tools are available at the FSSC with their accompanying documentation[2].

This Chapter will introduce a number of these tools and how they practically achieve the mathematics detailed in previous Chapters. Due to the relative freedom as to how to use these tools, a number of the practices recommended by the Fermi collaboration will also be discussed.

[1] Available at the FSSC website fermi.gsfc.nasa.gov/ssc/data/analysis/software/

[2] References and documentation on all FermiTools functions can be found at fermi.gsfc.nasa.gov/ssc/data/analysis/scitools/references.html

5.1 Types of Analysis

There are two types of analyses that are typically used in Fermi studies. These are the broadband fit, and the multiband fit. Their use is motivated by the needs of the research, however the multiband fit possesses certain advantages to the broadband in certain situations.

For data from a certain energy range (say 100 MeV to 100 GeV of typical studies), a broadband (or global) fit will assume that the source has a particular model across the entire range. Running a maximum likelihood analysis will find the best estimates to the normalisation and shape parameters of the model. One can then assess the quality of the fit in various ways.

On the other hand, a multiband (sometimes known as model-independent) fit does not make the assumption of a particular spectral shape for the entire energy range. Instead, the energy range is divided into a number of energy bins, in which a power-law model is fit individually in each bin. Power laws are both simple and ubiquitous spectra and so are a reasonable assumption to make. Typically there is little dependence on the slope of the power-law due to the bins being narrow, and the parameter is fixed to a usual value of $\Gamma = 2$. Though in some instances, the results from a broadband analysis can be used to define this fixed parameter or to provide an initial 'guess' for the fitting algorithm to start from. This leaves the normalisation parameter free to be fit during each analysis.

This can be an advantageous method when the model is not fully understood across the entire spectrum, or when deviations from an assumed model are to be studied. For example, in the Galactic Center region there are significant uncertainties in the diffuse background model and so using a multiband fit can ease the restrictions of assumptions in this complex model.

All Fermi data must be binned spatially for any analysis, while the energy bins are either 'binned' or 'unbinned'. In this study, due to the great number of detected counts, binned analyses are used. It is important to note that both these processes described above are still binned analyses, despite the fact that it appears only the multiband fit has energy bins. In the broadband fit, a binned analysis is performed across the entire range. While in the multiband fit, individual binned analyses are performed in each energy bin.

5.2 Data Structures

The FITS[3] file structure is used extensively throughout astronomy disciplines and as such it is used by the FermiTools program to hold various data and images. While the design of the FITS files and their 'Header and Data Units' will be omitted in this thesis, the practical uses of this design will now be discussed. The standard way in which FITS files are created ensures that the data can be easily accessed by FermiTools and by other third party programs.

Reconstructed data gathered by Fermi are held in FITS files with the designation 'FT1'. These event files hold a table of data with each row representing a detected signal, and each of the reconstructed characteristics are held in the various columns. These include the energy, direction, arrival time, etc.

The spacecraft file, with 'FT2' designation, holds the information of the position and pointing direction of the Fermi telescope in 30 second intervals within a table structure. The table also indicates the position relative to the Earth, and the condition of the data collected during the interval.

The output of the various FermiTool functions will be stored in FITS files as well. These do not carry the FT1 or FT2 file extensions, but will hold information in additional table and image extensions as required.

Sources that are modelled and to be fit using FermiTools are recorded in an XML data structure. However, for historical reasons during development, the XML structure is not unique across all FermiTool functions. The XML file required for simulating data, and the XML file required for the maximum likelihood method are similar but are not cross-compatible. The definitions available and way to structure each file can be found elsewhere[4]. However, both are organised into 'source libraries' in which individual sources are written in multi-line tags with various elements and attributes defining the spectrum and spatial descriptions. Parameters given in the likelihood XML file are given an element '`free`', that determines if the parameter is fixed (`free` = 0), or if it is to be fit by the algorithm (`free` = 1).

Below, a gamma ray source named 'Example Point Source' is defined for both XML files. The source has a power law model with an integrated flux of 0.0093 photons $m^{-2}s^{-1}$ between 30 MeV and 10000 MeV, with a spectral index of 1.6. The point source is in the direction of $(\alpha, \delta) = (128.7, -45.2)$. The code blocks illustrate the similarities

[3] fits.gsfc.nasa.gov/fits_home.html

[4] The definitions for `gtobssim` can be found at fermi.gsfc.nasa.gov/ssc/data/analysis/scitools/other_sources.html, and for `gtlike` at fermi.gsfc.nasa.gov/ssc/data/analysis/scitools/xml_model_defs.html#xmlModelDefinitions.

and differences when applied to both XML structures, and will be how they are used throughout this thesis.

```
<source name="Example Point Source" flux="0.0093">
   <spectrum escale="MeV">
      <particle name="gamma">
         <power_law emin="30" emax="10000" gamma="1.6"/>
      </particle>
      <celestial_dir ra="128.7" dec="-45.2"/>
   </spectrum>
</source>
```

CODE 5.1: Example of a Point Source XML definition for `gtobssim`.

```
<source name="Example Point Source" type="PointSource">
<spectrum type="PowerLaw2">
    <parameter free="1" max="100" min="1e-05" name="Integral"
                                                scale="1e-06" value="0.93"/>
    <parameter free="1" max="-1.0" min="-5.0" name="Index"
                                                scale="1.0" value="-1.6"/>
    <parameter free="0" max="10000" min="30" name="LowerLimit"
                                                scale="1.0" value="30"/>
    <parameter free="0" max="10000" min="30" name="UpperLimit"
                                                scale="1.0" value="10000"/>
</spectrum>
<spatialModel type="SkyDirFunction">
    <parameter free="0" max="360." min="-360." name="RA" scale="1.0" value="128.7"/>
    <parameter free="0" max="90." min="-90." name="DEC" scale="1.0" value="-45.2"/>
</spatialModel>
</source>
```

CODE 5.2: Example of a Point Source XML definition for `gtlike`. Due to the length of certain tags, these tags have been split over two lines.

5.3 Observation Simulation

Analysis is often performed on actual data measured by Fermi and made available on the FSSC servers. In studies, such as this, in which some level of consistency and ability to control the data is required for the validation of expected results, one must simulate a set of data. This is achieved in the FermiTools package with `gtobssim`.

Given a history of the observing telescope, its pointing direction and exposure over a given period, `gtobssim` will simulate the telescope collecting photons emitted by sources in a particular model. The faux detections are subject to the dispersion effects of the IRF of the telescope, and so what is simulated corresponds to the reconstructions of gamma ray events. As the simulations need to be randomised, a random number seed is also given to the function. A different seed number given will result in a different

simulation of events, while providing the same input and seed number will allow the recovery of the same simulation output.

The pointing history of the telescope is input to `gtobssim` as an FT2 file. Alternatively, a history can be generated on the fly or can be created by `gtorbsim` following simple orbits defined by a number of additional parameters. Throughout this study, the actual spacecraft file of the Fermi telescope will be used as it is simple to obtain and implement.

`gtobssim` is able to simulate gamma ray data through a cyclical process that repeats as required until the time period of interest has been met. Starting with all the sources defined in the XML file placed onto the celestial sphere around the telescope, the internal algorithm polls the sources in order to determine which will be the next source to emit a gamma ray. This is determined from the spectrum of the various sources. The algorithm will then compute the energy of the gamma ray and when the source will emit the signal.

In order for the telescope to detect this signal, `gtobssim` calculates the position and pointing direction and determines from the response functions the measured direction and energy of the gamma ray, if it is detected at all. This cycle is then repeated until the next source polled to emit a gamma ray, emits outside the duration of time being considered.

5.4 Data Selection

Generally, data taken from the FSSC servers or simulated by `gtobssim` extend beyond the ranges of energy, space, and time that are of interest to a study. Cuts must be made to the data in order to trim any excess to only what is relevant. This is achieved with both `gtselect` and `gtmktime`. `gtbin` is then used to create a binned domain in energy and space.

The `gtselect` tool reads the event files containing the reconstructed photon data and outputs another FITS file containing the filtered events. Most cuts are relatively straightforward, such as selecting energy ranges, a region of interest or acceptance cone around a certain direction, and the start and end times for the data. One can also make selections on the event class and event type, the classifications that describe the quality of reconstruction and probability of gamma ray detection. The final selection cut that can be made is of the zenith angle of a detected event. Due to the Earth limb acting as a source of gamma rays, events that arrived with a zenith angle greater than $\theta_Z \geq 90^\circ$ are typically poor candidates for extrasolar gamma rays. The recommended practice is to cut at 90°, but in some studies 105° is taken to include more photons in the selection. `gtselect` reads each row of the Table unit of the events file and checks to see if each

column value satisfies the selection cuts that have been made. If they do, they are copied over to the new FITS file to be stored.

The Fermi telescope is not always in data collecting mode, nor is it always collecting quality data. The `gtmktime` tool gives the ability to account for this and exclude periods when certain events have affected the quality. The tool works in two parts, the first is to create a sequence of Good Time Intervals (GTIs) in which the selection criteria are satisfied, and then to read the input events file to find events that were detected during these GTIs and add them to a new FITS file.

In each 30 second interval of the spacecraft pointing history, the `DATA_QUAL` and `LAT_CONFIG` quantities are given. `DATA_QUAL` is a flag that indicates the quality, a value of 0 being bad data, and 1 being good data. Negative values indicating that the IRFs do not describe the data correctly, such as during Solar flares and particle events. Values of 2 and 3 indicate that there may be a timing anomaly or that there are some bad parts. Similarly, if the LAT is configured for testing or non-standard data collecting, the `LAT_CONFIG` flag is set to 0 from the normal science configuration of 1. While not currently recommended, the `ROCK_ANGLE` can also be used to remove intervals in which the telescope has been orientated for viewing of a Target of Opportunity.

The selection criteria of `gtmktime` are expressed as a logical statement using functions such as the 'and', 'or', and 'not' terms. In this syntax, the recommended filter is given as `(DATA_QUAL>0) && (LAT_CONFIG==1)`.

The final step is to create a binned counts cube of the data expressed as a series of images with each pixel representing the number of counts that occur in that specific bin. The spatial domain of the counts cube is a rectangular region (in right ascension, declination space) that must fit within the region selected for by `gtselect`. Typically, the largest square region within the circular acceptance cone is used. Thus, the sides of the square are a factor of $\sqrt{2}$ larger than the radius of the acceptance cone. Defining a pixel size then allows the transformation of this length into a number of pixels. When running the `gtbin` function, the center of the rectangular region must also be given in either the Celestial or Galactic coordinates, and the particular projection method to be used.

The energy axis of the counts cube is typically binned in equally sized logarithmic bins, though the option is given to choose linearly spaced bins or even to import a file with a custom defined binning arrangement. A typical gamma ray source spectrum spans several decades, and so it is often natural to use the logarithmic bins to account for this. For a narrow study of a particular energy band, a linear binning may be more

appropriate. The upper and lower bounds of the entire energy range, and the number of uniform energy bins, are sufficient for `gtbin` to construct the counts cube energy axis.

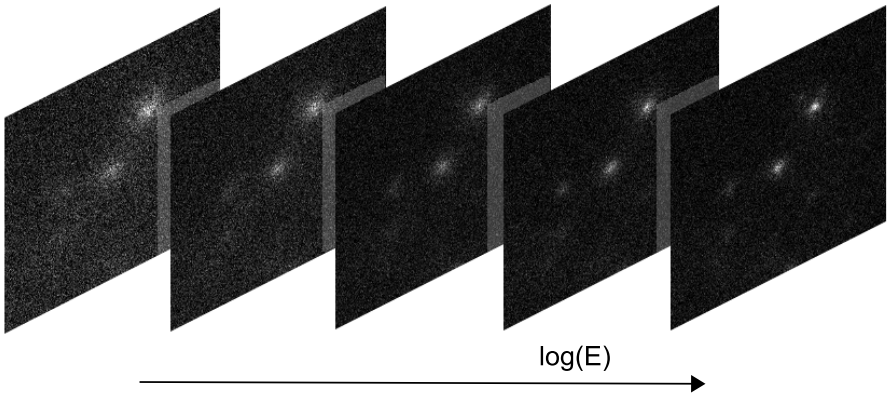


FIGURE 5.1: First five slices of a counts cube generated from data around $(\alpha, \delta) = (193.98, -5.82)$ for the first two years of data collection. The bins are centered on 112 MeV, 141 MeV, 178 MeV, 224 MeV, and 282 MeV. A transparent square has been added in each corner in order to show the different layers stacked.

5.5 Exposure Cubes

To ease computational workloads during the model fitting stage, a number of model-independent quantities are pre-computed. If these are not pre-computed before envoking the `gtlike` function, these quantities would have to be calculated several times as the algorithm runs. As was shown in Chapter 3 in the derivation of the counts statistic the exposure, and implicitly the livetime, could be factorised out of the integration, allowing them to be calculated once and be reused.

The livetime cube represents the total amount of time that the Fermi telescope has spent observing at a particular direction and inclination while taking data. This is achieved using the `gtltcube` function, which takes the pointing history of the telescope and bins the accumulated time spent at each spatial position and at each inclination angle. The spatial bins are binned in equal surface area divisions of a sphere using a HEALPIX grid[5].The FSSC suggests using a binning size of $\Delta\cos(\theta) = 0.025$ and a HEALPIX pixel size of 1° in typical studies.

The number of counts expected at a certain spatial position and energy range, depend on the amount of exposure taken for those parameters. The exposure, the product of the area and time, depends only on the IRFs and the livetime as shown in Equation 3.17.

[5]HEALPIX stands for Hierarchical Equal Area isoLatitude Pixelization. http://healpix.jpl.nasa.gov/

FermiTools creates a three-dimensional cube of the exposure using `gtexpcube2` that is binned in spatial position and energy, the same way `gtselect` binned the original data.

While the sources within the Region of Interest are most likely to contribute to the counts, sources outside this region can also contribute. In this case the exposure must be calculated for a wider region than was defined for `gtbin`, while keeping the same central coordinates, pixel size, and energy binning. The recommended buffer region depends on the energy ranges considered, though is typically a further 10° to 20°. In some instances, especially with diffuse extended sources, it may be easier to extend the buffer region to include the entire $360° \times 180°$ sky. Though this can increase computation time somewhat.

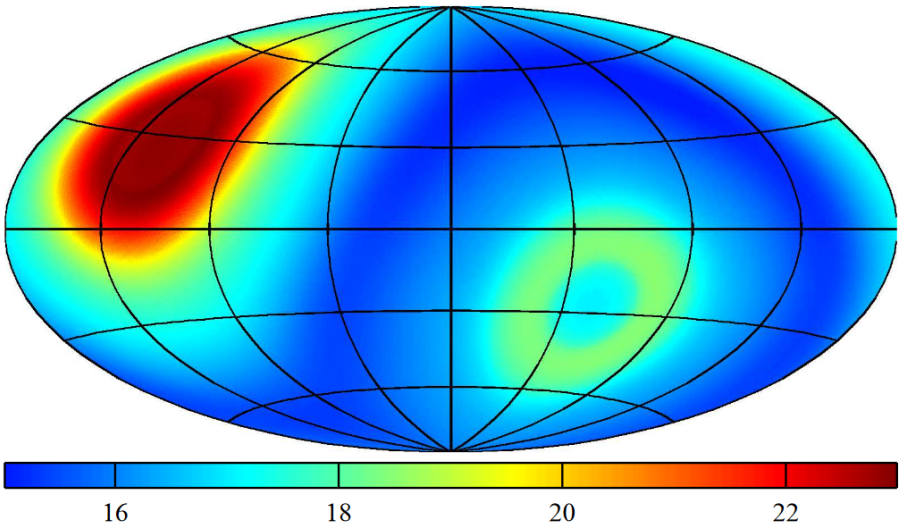


FIGURE 5.2: Exposure at 1 GeV for the 4 years of Fermi data considered in the 3FGL. It is expressed in equivalent on-axis observing time (Ms) and shown in Galactic Coordinates of a Hammer-Aitoff projection. Figure obtained from [8].

The final pre-computed quantity is the source map. This FermiTools function, `gtsrcmaps`, takes the model of gamma ray sources and convolves them with the IRFs to produce a model counts cube that is used in the likelihood fitting. The FITS file produced contains an extension for each source included in the model, containing just the counts cube of that source. The individual counts maps are binned, as before, in energy and spatial position per the geometry given for `gtbin`.

5.6 Likelihood Fitting

At this point, the necessary files have been constructed that allow the likelihood maximising algorithm to proceed and fit the model to the data. In the FermiTools program this function is called `gtlike`.

Following the discussion in Chapter 4, `gtlike` must solve the non-linear functions that achieve the maximum of the likelihood function for the Poisson distribution. Beginning from an initial guess as to the model parameters given in the XML file, an algorithm will compute the log-likelihood function for these particular parameter choices. It will then determine a direction in parameter space that will improve the likelihood function, and change the initial guess to the updated estimate $\boldsymbol{\theta} \rightarrow \boldsymbol{\theta} + d\boldsymbol{\theta}$. The direction is found by estimating the derivatives of the likelihood function with respect to the various parameters that are being fitted. The iterative process will then calculate the likelihood function again and continue to update the parameter estimates until convergence to a certain tolerance is reached (or in poor circumstances, a failure to converge).

Depending on the circumstances, FermiTools offers a number of optimising algorithms to perform this task. Each optimiser makes a play-off between speed and precision. Throughout this study, the `NEWMINUIT` optimiser[6] will be used. This algorithm, while slower in convergence than other options, provides a more detailed mapping of the likelihoods dependence on parameters near the maximum. From this mapping, the uncertainties of the fitted parameters can be derived from the Hessian matrix evaluated at the MLE.

The following segment of XML code, shows the XML output from `gtlike`. The file is very similar to the input version, though the output contains the `error` attribute for each parameter that had the `free` flag set to 1.

```
<source name="Example Point Source" type="PointSource">
<spectrum type="PowerLaw2">
    <parameter error="0.3" free="1" max="100" min="1e-05" name="Integral"
                                                    scale="1e-06" value="0.97"/>
    <parameter error="0.2" free="1" max="-1.0" min="-5.0" name="Index"
                                                    scale="1.0" value="-1.5"/>
    <parameter free="0" max="10000" min="30" name="LowerLimit"
                                                    scale="1.0" value="30"/>
    <parameter free="0" max="10000" min="30" name="UpperLimit"
                                                    scale="1.0" value="10000"/>
</spectrum>
<spatialModel type="SkyDirFunction">
    <parameter free="0" max="360." min="-360." name="RA" scale="1.0" value="128.7"/>
    <parameter free="0" max="90." min="-90." name="DEC" scale="1.0" value="-45.2"/>
</spatialModel>
</source>
```

CODE 5.3: Example of a Point Source XML definition after being fit by `gtlike`. Due to the length of certain tags, these tags have been split over two lines.

[6] The user guide, and source code for MINUIT and related optimisers can be found at lcgapp.cern.ch/project/cls/work-packages/mathlibs/minuit/doc/doc.html

Chapter 6

Statistical Analysis

The maximum likelihood method was introduced in Chapter 4 as a way of determining the best estimates to the parameters of a particular model for some data set. The likelihood function was presented as a statistic that measured the probability of observing the data given a model and choice of parameters. The properties of the MLE were discussed, highlighting the usefulness of this particular inference theory.

It was mentioned that the likelihood function can only be compared between a single data set, and not across different sets of data. That is, the likelihood for some data set under different models or different choices of parameters can all be compared against each other. This motivates the discussion of hypothesis testing and formalising a procedure to choose between models or choices of parameters.

The Neyman-Pearson lemma states that for simple hypotheses, the ratio of likelihoods under the different hypotheses gives the most powerful test in accepting or rejecting the choices at a certain significance level. One can then expand upon this lemma in order to consider composite hypotheses and, of interest in this study, nested models.

In the limit of large sample size, the likelihood ratio follows a χ^2 distribution with degrees of freedom equal to the difference in dimensionality of the nested models. This result, first derived by Wilks in 1938, provides a useful property in considering the goodness of fit of a model and allows the calculation of the p-value under reasonable assumptions.

This Chapter will present a derivation of the asymptotic distribution for the likelihood ratio in a multiband analysis. This will be applied to the hypothesis test of determining if deviations in a background model are statistically significant or not to warrant the addition of a point source to the model. The distribution of the flux of the point source will also be discussed.

6.1 Hypothesis Testing

A hypothesis, in the statistical sense of the word, concerns the choices of parameters of a model or probability distribution that generates some form of measurable data. Often different hypotheses are suggested as the explanation of how some particular data set arose and so one must find some way to decide between them. A hypothesis test formalises this notion and presents the procedure taken to determine whether one hypothesis of a model should be accepted over another hypothesis. The null hypothesis, H_0, is defined as the hypothesis in which the observed data deviates from the model due to random chance. The alternative hypothesis, H_1, is the explanation that the deviations arise due to a non-random process. The structure of a hypothesis test allows the hands-free approach to choosing between models, providing a mathematical approach to deciding when their is sufficient or insufficient evidence to reject the null hypothesis.

It is important to note that a hypothesis test does not say when a null hypothesis should be accepted. It merely states that the test has or has not failed to reject it.

When all the parameters of the model are specified, the hypothesis is referred to as a simple hypothesis. For example, data taken from a Poissonian distribution could offer two simple hypotheses that the parameter takes on $\lambda = 1$ or $\lambda = 2$. However, not all cases are like this. Composite hypotheses on the other hand are cases when not all the model parameters have been specified. Following the previous example, one could test the data to determine if $0 \leq \lambda < 1$ or if $\lambda \geq 1$.

Ultimately, a hypothesis test comes down to forming a rule on when to reject the null hypothesis. A test statistic quantity that depends on both hypotheses and the data, is used to define a threshold value. If the observed test statistic is greater or less than this threshold, then there is sufficient evidence to reject or not reject one of the hypotheses.

Let a set of observations $\mathbf{x}$ be described by a model hypothesis denoted H_0, and that an alternative hypothesis, H_1, to the model has been suggested. The two hypotheses have a probability distribution function for the test statistic t, given by $g(t|H_0)$ and $g(t|H_1)$. In order to reject or fail to reject the null hypothesis, a critical region is defined for the test statistic. If t is calculated to be within the critical region, then H_0 is rejected. The probability of observing the test statistic in the critical region is denoted α, the significance level.

For the purposes of the analysis presented, the critical region is defined as the range of t values above a certain cut-off value t_{cut}. Thus, the significance level is given by

$$\alpha = \int_{t_{cut}}^{\infty} g(t|H_0)dt \tag{6.1}$$

This represents the Type I error, or the probability that H_0 is rejected if H_0 is indeed the true model.

The Type II error occurs when H_0 is accepted but the true model was actually H_1. This is the power of the test, $1 - \beta$, where

$$\beta = \int_{-\infty}^{t_{cut}} g(t|H_1)dt \ . \tag{6.2}$$

6.2 Neyman-Pearson Lemma

The above frame work laid out the methodology of hypothesis testing, however it did not give any indication to how one should choose a test statistic or decide on the accompanying critical region. Typically, for some significance level, one wants to achieve the highest power statistic, minmising the Type II errors that can occur. If both the null and alternate hypotheses are simple, then the Neyman-Pearson Lemma gives the most powerful test possible at significance level α [48].

The test statistic in this test is defined as the likelihood ratio, $\Lambda(\mathbf{x})$.

$$\Lambda(\mathbf{x}) = \frac{L(H_0|\mathbf{x})}{L(H_1|\mathbf{x})} \ . \tag{6.3}$$

If $\Lambda(\mathbf{x})$ is calculated to be less than some threshold value η, then the null hypothesis must be rejected. The threshold value is strongly tied to the significance level, α, according to

$$\alpha = P(\Lambda(\mathbf{x}) \leq \eta|H_0) \ . \tag{6.4}$$

It is common to first choose a significance level, say 5%, and then deduce the value of η from the above equation.

6.3 Wilks Theorem

The Neyman-Pearson lemma establishes that the likelihood ratio for simple hypothesis is the most powerful statistic to use. The concept can be further generalised to nested composite hypotheses by considering the supremum of the likelihood function under each hypothesis.

$$\Lambda(\mathbf{x}) = \frac{\sup\{L(\theta|\mathbf{x}) : \theta \in \Theta_0\}}{\sup\{L(\theta|\mathbf{x}) : \theta \in \Theta_1\}} \tag{6.5}$$

The model of the probability distribution contains k parameters from the parameter space $\Theta_1 = \{\theta_1, \theta_2, ..., \theta_k\}$. In the alternative hypothesis, all these parameters are unspecified. While the first j of these are specified at certain values in the null model. The more general model can be reduced to this null model through the choices of these parameters. The parameter space $\Theta_0 = \{\theta_1^0, \theta_2^0, ..., \theta_j^0, \theta_{j+1}, ..., \theta_k\}$ is thus a subset of Θ_1, where the 0 superscript indicates the taking on of a specific value. The difference in dimensionality between these two parameter spaces is $n = k - j$.

It has been shown by defining a test statistic TS as $-2\ln(\Lambda(\mathbf{x}))$, that the asymptotic distribution of TS follows a χ^2 distribution with n degrees of freedom as defined above [49].

$$-2\ln(\Lambda(\mathbf{x})) \sim \chi_n^2 \tag{6.6}$$

This result, known as Wilks theorem, relies on the assumption of the regularity conditions of the maximum likelihood method that were discussed in Chapter 4. It also assumes that the true data generating distribution of $\mathbf{x}$ is the null hypothesis.

As the χ^2 distribution is well studied, Wilks theorem provides a convenient result allowing the calculation of the p-value with relative ease. It is used extensively in Fermi studies and gamma ray astronomy.

The regularity conditions required that the parameters of the model be interior points of the parameter space. This is sometimes broken in practical situations. The particle flux of a gamma ray source can not be a negative value, a sink of gamma ray emissions. In effect, the lower bound of the flux parameter must be 0.

Point source detection relies on this when trying to determine if the data observed is explained best by random deviations from a background model, or if the deviations come from the presence of a point source. When the true underlying model does not have a point source present, the fit will converge to a flux value of 0. While when it has found a source, it will converge above this value. This can be seen as the null hypothesis (a background source) being a special nested case of the alternative hypothesis (a background source and a point source) with the normalisation of the point source specified as 0.

Wilks theorem does not account for this result, and the asymptotic distribution does not follow the simple χ^2 relation. However, it can be shown that the Test Statistic is still of the quadratic form and instead follows a mixture distribution [50][51].

$$TS \sim \frac{1}{2}\left(\chi_0^2 + \chi_1^2\right) \tag{6.7}$$

This TS is distributed such that half of the time it will be zero, and the other half being drawn from a $\frac{1}{2}\chi^2_1$ distribution.

It should be noted that the χ^2 distribution is usually defined to have a positive integer number of degrees of freedom, and as such a χ^2_0 distribution would be meaningless. However, this notation is used in place of a Dirac delta function distribution where $P(X = x) = \delta(x)$. In the following sections, when the sum of χ^2 terms is taken, it is convenient to use this notation.

This result has previously been empirically shown through the Monte Carlo simulation of gamma ray data for EGRET. A background model is simulated and then fit with and without the assumptions that the deviations are explained by an additional point source. This result is reproduced below. The simulations illustrate that the asymptotic distribution is not predicted by Wilks theorem, and instead follows the $\frac{1}{2}\chi^2_1$ distribution.

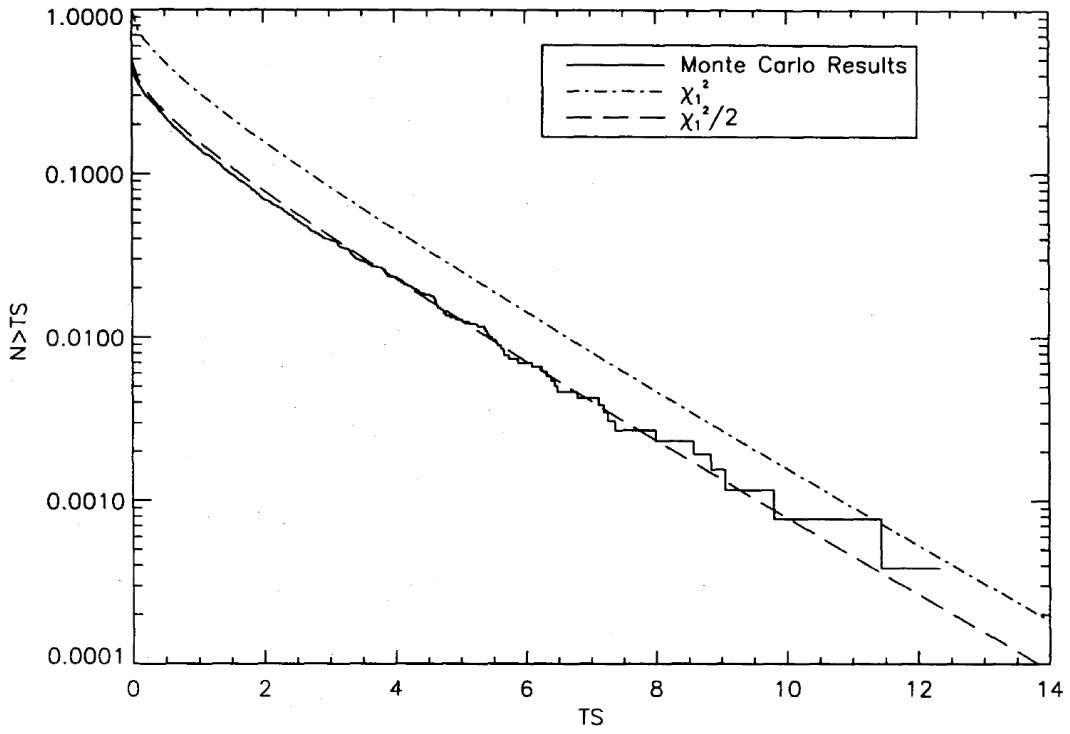


FIGURE 6.1: The Cumulative Distribution of TS following the $\frac{1}{2}\chi^2_1$ distribution. Several thousand simulations of a background source were taken and fit both with and without a point source. The TS results of these simulations are shown in the unbroken line, and can be seen to clearly follow the expected distribution and not the χ^2_1 distribution predicted by Wilks Theorem. Figure obtained from [12].

6.4 Multiband Distribution

The previous results have considered the broadband style of analysis. In general, if multiple bins are considered then one must multiply the likelihoods of each bin together. Due

to the properties of the log transformation, adding the log-likelihoods is an equivalent process and more relevant to the definition of the test statistic.

For N bins, the total test statistic is calculated by the sum

$$\begin{aligned} TS &= \sum_{i=1}^{N} TS_i \\ &= \sum_{i=1}^{N} -2\left(\ln(L_i(\mathbf{x}|H_0)) - \ln(L_i(\mathbf{x}|H_1))\right) \ . \end{aligned} \tag{6.8}$$

In the multiband analysis that is being considered in this study, parameters are fit for the normalisation of a power law model in each bin independently of each other. If the null hypothesis has a specified parameter on the boundary of the parameter space of the alternative hypothesis, then each bin will have a TS_i distributed as

$$TS_i \sim \frac{1}{2}\left(\chi_0^2 + \chi_1^2\right) \ . \tag{6.9}$$

It will now be derived that the sum of these TS_i values follows a mixture distribution of χ^2 distributions weighted by binomial coefficients.

$$TS \sim \frac{1}{2^N} \sum_{i=0}^{N} \binom{N}{i} \chi_i^2 \ . \tag{6.10}$$

6.4.1 Moment Generating Function Theorems

Before the derivation of the Test Statistic is carried out, two theorems must first be shown as they will be used in the full derivation.

Theorem 6.1. *Let X be a continuous random variable with a finite mixture distribution given by*

$$f(x) = \sum_{i=0}^{N} a_i f_i(x) \tag{6.11}$$

where a_i are positive real coefficeints, and $f_i(x)$ are probability density functions. Then the moment generating function of X, $M_X(t)$, is given by

$$M_X(t) = \sum_{i=0}^{N} a_i M_i(t) \tag{6.12}$$

where $M_i(t)$ is the moment generating function corresponding to $f_i(x)$.

Proof. This can be shown through the definition of the moment generating function and sufficient rearrangement of the weighted terms.

$$\begin{aligned} M_X(t) = \mathrm{E}[e^{Xt}] &= \int_{-\infty}^{+\infty} e^{xt} f(x) dx \\ &= \int_{-\infty}^{+\infty} e^{xt} \left(\sum_{i=0}^{N} a_i f_i(x) \right) dx \end{aligned}$$

The integral of a sum of terms is equal to the sum of integrals, allowing the summation term to come outside of the integral.

$$M_X(t) = \sum_{i=0}^{N} \int_{-\infty}^{+\infty} e^{xt} a_i f_i(x) dx$$

Taking the constant weight value outside the integral as well, one is left with the moment generating function definition for each component of the mixture.

$$\begin{aligned} M_X(t) &= \sum_{i=0}^{N} a_i \int_{-\infty}^{+\infty} e^{xt} f_i(x) dx \\ &= \sum_{i=0}^{N} a_i M_i(t) \end{aligned}$$

□

Theorem 6.2. *Let X be a continuous random variable, being the sum of N independent and identically distributed random variables X_i, $X = \sum_{i=0}^{N} X_i$. As each X_i is an iid variable, each have the same moment generating function, given by M_{X_0}. The moment generating function of X is*

$$M_X(t) = \prod_{i=0}^{N} M_{X_i}(t) \ . \tag{6.13}$$

Proof. This theorem will be shown through manipulation of the expectation value definition of the moment generating function.

$$\begin{aligned} M_X(t) &= \mathrm{E}[e^{Xt}] \\ &= \mathrm{E}\left[e^{\left(\sum_{i=0}^{N} X_i\right)t} \right] \\ &= \mathrm{E}\left[e^{\left(\sum_{i=0}^{N} X_i t\right)} \right] \end{aligned}$$

The sum of exponential terms can be split into a product of individual exponential terms. The resulting product notation can be taken outside of the expectation value, as

each X_i are independent of each other.

$$\begin{aligned} M_X(t) &= \mathrm{E}\left[\prod_{i=0}^{N} e^{X_i t}\right] \\ &= \prod_{i=0}^{N} \mathrm{E}\left[e^{X_i t}\right] \end{aligned}$$

This leaves a product of the moment generating functions for each X_i, however as they are all identically distributed, they each possess the same generating function.

$$\begin{aligned} M_X(t) &= \prod_{i=0}^{N} M_{X_i}(t) \\ &= [M_{X_0}(t)]^N \end{aligned}$$

□

6.4.2 TS Distribution Derivation

In a one bin scenario, the distribution of the test statistic takes on the form of a mixed distribution of a χ^2 distribution with one degree of freedom and a Dirac delta function.

$$TS_0 \sim \frac{1}{2}\left(\chi_0^2 + \chi_1^2\right) \tag{6.14}$$

Thus, the probability density function of TS_0 is given by

$$f(x) = \frac{1}{2}\left(f_{\chi_0^2}(x) + f_{\chi_1^2}(x)\right) \tag{6.15}$$

where $f_{\chi_0^2}(x)$ and $f_{\chi_1^2}(x)$ are the probability density functions of their respective distributions.

Using Theorem 6.1, the moment generating function of TS_1 is given by the weighted sum of the moment generating functions of each of the distributions.

$$M_{TS_0}(t) = \frac{1}{2}\left(1 + (1-2t)^{-\frac{1}{2}}\right) \tag{6.16}$$

The derivation of the moment generating functions for these distributions will be omitted as they are standard results shown frequently elsewhere [52].

Consider N bins that are fit according to the prescription detailed above. The sum of likelihood ratio statistics $TS = \sum_{i=0}^{N} TS_i$ will have a moment generating function,

according to Theorem 6.2, given by

$$M_{TS}(t) = \left[M_{TS_0}(t)\right]^N \tag{6.17}$$

where $M_{TS_0}(t) = \frac{1}{2}\left(1 + (1-2t)^{-\frac{1}{2}}\right)$ is the moment generating function of the distribution of an individual bin.

One can then expand the polynomial using the binomial theorem to arrive at an expression in terms of the moment generating functions of χ_i^2.

$$\begin{aligned} M_{TS}(t) &= \left[\frac{1}{2}\left(1 + (1-2t)^{-\frac{1}{2}}\right)\right]^N \\ &= \frac{1}{2^N}\sum_{i=0}^{N}\binom{N}{i} 1^{N-i}\,(1-2t)^{-i/2} \\ &= \frac{1}{2^N}\sum_{i=0}^{N}\binom{N}{i}\,(1-2t)^{-i/2} \\ &= \sum_{i=0}^{N}\frac{1}{2^N}\binom{N}{i} M_{\chi_i^2}(t) \end{aligned} \tag{6.18}$$

The final expression of the above equation, expresses the moment generating function of the TS distribution as a weighted sum of the moment generating functions of χ_i^2 distributions. This is of the form that was derived in Theorem 6.1, and so one can apply the theorem in reverse to derive the distribution. Thus,

$$X \sim \frac{1}{2^N}\sum_{i=0}^{N}\binom{N}{i}\chi_i^2 \ . \tag{6.19}$$

6.5 Point Source Normalisation

The properties of the MLE included an asymptotic normal distribution of the fitted parameters around their true values. When considering the situation described above in determining whether fluctuations in a background source are statistically significant to be fit as an additional point source, the normalisation of the point source will follow a normal distribution. As the data is known to come from a background only model, the true value of the point source normalisation must be 0. Thus, the normal distribution will be centered on $\mu = 0$ as well with variance σ_N^2.

However, the normalisation is bounded below by 0. A gamma ray source can not act as a sink of flux, and so only positive values make physical sense to consider. In the absence of

this restriction, the normal distribution would allow these negative normalisation values to be drawn from. With the restriction in place, all negative fluxes are forced to take the value 0.

The distribution of the normalisation no longer follows a normal distribution, and instead follows a piece-wise function given by,

$$X \sim \begin{cases} 0 & x < 0 \\ \frac{1}{2}\delta(x) + N(0, \sigma_N^2) & 0 \leq x \end{cases} . \tag{6.20}$$

This section will derive the expectation and variance properties of this distribution.

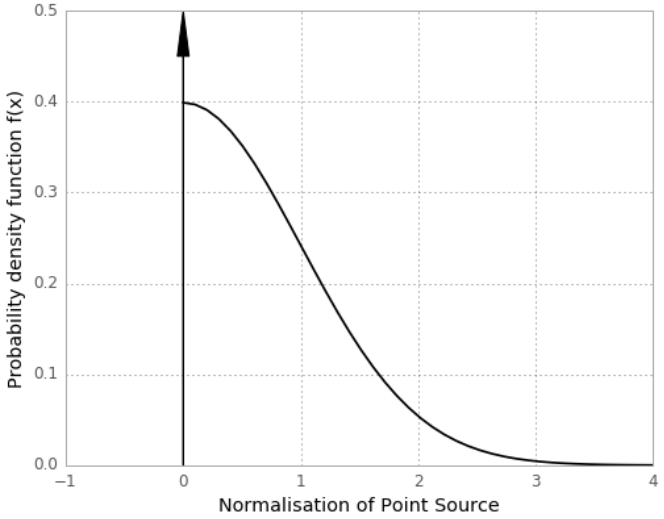


FIGURE 6.2: Distribution of the point source normalisation. The normal distribution has a mean $\mu = 0$ and variance $\sigma_N^2 = 1$.

The expected value of this distribution can be derived from first principles as such. The total integral across the entire noramlisation range can be broken into three parts, corresponding to the three features of the distribution. The form of the Dirac delta function can be taken as any of the definitions that in the limit $\epsilon \to 0$, recovers the infinitesimally thin peak. The definition is not shown in the below derivation fore brevity,

as it is a standard result.

$$\begin{aligned}
\mathrm{E}[X] &= \int_{-\infty}^{+\infty} xf(x)dx \\
&= \lim_{\epsilon\to 0^+} \int_{-\infty}^{-\epsilon} xf(x)dx + \int_{-\epsilon}^{+\epsilon} xf(x)dx + \int_{+\epsilon}^{+\infty} xf(x)dx \\
&= \lim_{\epsilon\to 0^+} \int_{-\epsilon}^{+\epsilon} x\frac{1}{2}\delta(x)dx + \int_{+\epsilon}^{+\infty} x\frac{1}{\sqrt{2\pi\sigma_N^2}} e^{-\frac{x^2}{2\sigma_N^2}} dx \\
&= \int_{0}^{+\infty} x\frac{1}{\sqrt{2\pi\sigma_N^2}} e^{-\frac{x^2}{2\sigma_N^2}} dx \\
&= \frac{\sigma_N}{\sqrt{2\pi}}
\end{aligned} \tag{6.21}$$

Similarly, one can derive the variance of the distribution using the definition $VAR[X] = E[X^2] - E[X]^2$. The expectation value of the X^2 distribution can be achieved in the same as before by splitting up the integral over the three features,

$$\begin{aligned}
\mathrm{E}[X^2] &= \int_{-\infty}^{+\infty} x^2 f(x)dx \\
&= \lim_{\epsilon\to 0} \int_{-\infty}^{-\epsilon} x^2 f(x)dx + \int_{-\epsilon}^{+\epsilon} x^2 f(x)dx + \int_{+\epsilon}^{+\infty} x^2 f(x)dx \\
&= \lim_{\epsilon\to 0} \int_{-\infty}^{-\epsilon} x^2 0dx + \int_{-\epsilon}^{+\epsilon} x^2\frac{1}{2}\delta(x)dx + \int_{+\epsilon}^{+\infty} x^2\frac{1}{\sqrt{(2\pi\sigma_N^2)}} e^{-\frac{x^2}{2\sigma_N^2}} dx \\
&= \int_{0}^{+\infty} x^2\frac{1}{\sqrt{2\pi\sigma_N^2}} e^{-\frac{x^2}{2\sigma_N^2}} dx \\
&= \frac{\sigma_N^2}{2} \ .
\end{aligned} \tag{6.22}$$

Combining the two results, one derives the variance of the distribution,

$$\begin{aligned}
\mathrm{VAR}[X] &= \mathrm{E}[X^2] - \mathrm{E}[X]^2 \\
&= \frac{\sigma_N^2}{2} - \frac{\sigma_N^2}{2\pi} \\
&= \frac{1}{2}\left(1 - \frac{1}{\pi}\right)\sigma_N^2 \ .
\end{aligned} \tag{6.23}$$

Chapter 7

Results

This Chapter will present the results of a Monte Carlo analysis of the sampling distribution of the likelihood ratio test statistic using a multiband analysis. 1000 simulations of an isotropic background model will be produced by `gtobssim` and through the various FermiTool functions of Chapter 5 will be fit by the maximum likelihood algorithm `gtlike`. Two hypotheses will be tested to model the simulated data. The results will be presented for $N = 2, 3, 4,$ and 5 bin multiband analyses.

The null hypothesis is that random chance has caused the fluctuations from the background model, while the alternative hypothesis is that there is statistically significant evidence of a point source in the data as well as a background source. It can be seen that the null hypothesis is a nested model of the alternative hypothesis, in which the normalisation of the test source has been set to zero. As the null distribution is known to be the correct model of the data, the derivations presented in the previous Chapter show that the likelihood ratio of the two hypothesis

$$TS = \sum_{i=1}^{N} -2\left(\ln(L_i(\mathbf{x}|H_0)) - \ln(L_i(\mathbf{x}|H_1))\right) \ , \tag{7.1}$$

will follow a mixed probability distribution

$$TS \sim \frac{1}{2^N} \sum_{i=0}^{N} \binom{N}{i} \chi_i^2 \ . \tag{7.2}$$

In keeping with the style of previous literature[12], the TS distributions will be graphed in terms of the survival function,

$$P(\mathrm{TS}_{\mathrm{obs}} \geq \mathrm{TS}) = 1 - \mathrm{CDF}(\mathrm{TS} \mid \mathrm{TS}_{\mathrm{obs}}) \ . \tag{7.3}$$

7.1 Data Simulation, Selection, and Fitting

1000 simulations were taken by `gtobssim` for a seven day observation beginning the 4th August 2008 at 2:43pm. The spacecraft file used to provide the pointing history of the Fermi telescope was the actual spacecraft file of Fermi used during this period, and not created by `gtorbsim`. The background source was modelled as an isotropic all-sky source with a power law spectrum. This spectrum has an integrated flux of 1000 photons $m^{-2}s^{-1}$ between 20 MeV and $2 \cdot 10^5$ MeV. For simplicity, the spectral index of the background was kept at $\Gamma = 2$.

```
<source_library title="Monte Carlo Background Simulation">
  <source name="Isotropic Background">
    <spectrum escale="MeV">
      <SpectrumClass name="Isotropic" params="1000, 2.0, 20, 2e5"/>
      <use_spectrum frame="galaxy"/>
    </spectrum>
  </source>
</source_library>
```

Code 7.1: XML source library used for the simulation of the isotropic background.

The energy range was split into N logarithmically spaced bands, and a binned analysis completed inside each band. The `SOURCE` class and Front/Back event types were used, as per the recommendations by the FSSC. `gtselect` was then used to select an acceptance cone of radius 10.7° around the (0,0) direction in Galactic coordinates. The zenith angle parameter was cut at the 90° angle. Likewise, `gtmktime` identified GTIs and selected for events that satisfied the `(DATA_QUAL>0) && (LAT_CONFIG==1)` filter expression.

An acceptance cone of size 10.7° allows a square with sides 15° long to be inscribed within. A pixel size of 0.1°/pix was used, defining a 150 pix by 150 pix binned domain in a Plate Carree projection. Within each band, energy was binned in 5 log-spaced bins.

The FSSC recommended practices were used to create the livetime and exposure cubes with `gtltcube` and `gtexpcube2` respectively. Two source maps were calculated by `gtsrcmaps` for the two models that will be fit.

The XML definitions, as required for `gtlike`, are shown below in Code 7.2. As the null hypothesis is simply the alternative hypothesis without the point source, only the alternative hypothesis XML file is shown for brevity. The point source is given a fixed spectral index of $\Gamma = 2$, like the isotropic background source. Only the normalisation parameters are flagged with the `free` flag in order to find their maximum likelihood estimates.

```
<source_library title="Alternative Hypothesis for likelihood fit">
<source name="Isotropic Background" type="DiffuseSource">
  <spectrum apply_edisp="false" type="PowerLaw2">
    <parameter free="1" max="10000" min="0" name="Integral" scale="1e-06"
                                                              value="1.0"/>
    <parameter free="0" max="-1" min="-5" name="Index" scale="1.0" value="-2.0"/>
    <parameter free="0" max="200000.0" min="20.0" name="LowerLimit"
                                                    scale="1.0" value="2e1"/>
    <parameter free="0" max="200000.0" min="20.0" name="UpperLimit"
                                                    scale="1.0" value="2e5"/>
  </spectrum>
  <spatialModel type="ConstantValue">
    <parameter free="0" max="10.0" min="0.0" name="Value" scale="1.0" value="1.0"/>
  </spatialModel>
</source>
<source name="Point Source" type="PointSource">
  <spectrum type="PowerLaw2">
    <parameter free="1" max="1000.0" min="0" name="Integral" scale="1e-06"
                                                              value="0"/>
    <parameter free="0" max="-1.0" min="-5.0" name="Index" scale="1.0"
                                                            value="-2.0"/>
    <parameter free="0" max="200000.0" min="20.0" name="LowerLimit"
                                                    scale="1.0" value="20.0"/>
    <parameter free="0" max="200000.0" min="20.0" name="UpperLimit"
                                                    scale="1.0" value="2e5"/>
  </spectrum>
  <spatialModel type="SkyDirFunction">
    <parameter free="0" max="360.0" min="-360.0" name="RA" scale="1.0" value="266.4"/>
    <parameter free="0" max="90.0" min="-90.0" name="DEC" scale="1.0" value="-28.9"/>
  </spatialModel>
</source>
</source_library>
```

CODE 7.2: XML source library used for the maximum likelihood fitting of the isotropic background source and a point source, according to the alternative hypothesis. Due to the lengths of certain tags, they have been split to two lines.

7.2 Results

Upon fitting, the TS values are found by Equation 7.1 and are graphed as a normalised histogram of survival function. In the graphs that have been produced in this Chapter, the histogram values are shown as the black stepped line and the expected analytic distribution from equation 7.2 is shown in red. In order to determine the goodness of fit between the simulated and analytic results, a 95% confidence interval is shown in green about the histogram line.

In each bin of the histogram, the probability of being above the TS value follows a binomial distribution. This is because there are only two outcomes, being above or

below the TS value, and that a fixed number of independent trials were carried out, in this case 1000 simulations. The normal approximation to the confidence interval does not necessarily hold for these cases, and so the Wilson score interval is used as it is known to be a good improvement.

The normalisation parameter for the point source in the alternative hypothesis is also presented as a histogram (the blue bars). Due to the nature of the parameters distribution, there is a large peak at $N = 0$ that would otherwise affect the aspect ratio of the graphs. A vertical axis cut allows the height of the peak to be seen, while also allowing the other histogram bins to be seen clearly.

Equation 6.21 derived the relationship between the expectation value of the normalisation parameter and the variance of the underlying normal distribution. The mean of the normalisation values is calculated, allowing the probability density function to be uniquely determined and shown as the dashed orange line on these plots.

7.3 Discussion

The results show that for each multiband analysis, the empirical distribution clearly matches the analytical function derived in the previous Chapter. It can be seen that the expected distribution falls within the 95% confidence interval, almost surely across the entire TS range. This Monte Carlo simulation validates the expectation of the TS distribution.

The normalisation also has been shown to follow the distribution discussed in the previous Chapter. A clear spike of the histogram at the origin indicates that a number of simulations determined there to be no point source present and assigned a normalisation value of zero. The rest of the distribution closely matches the normal distribution in the positive parameter region.

7.3.1 2 Bins

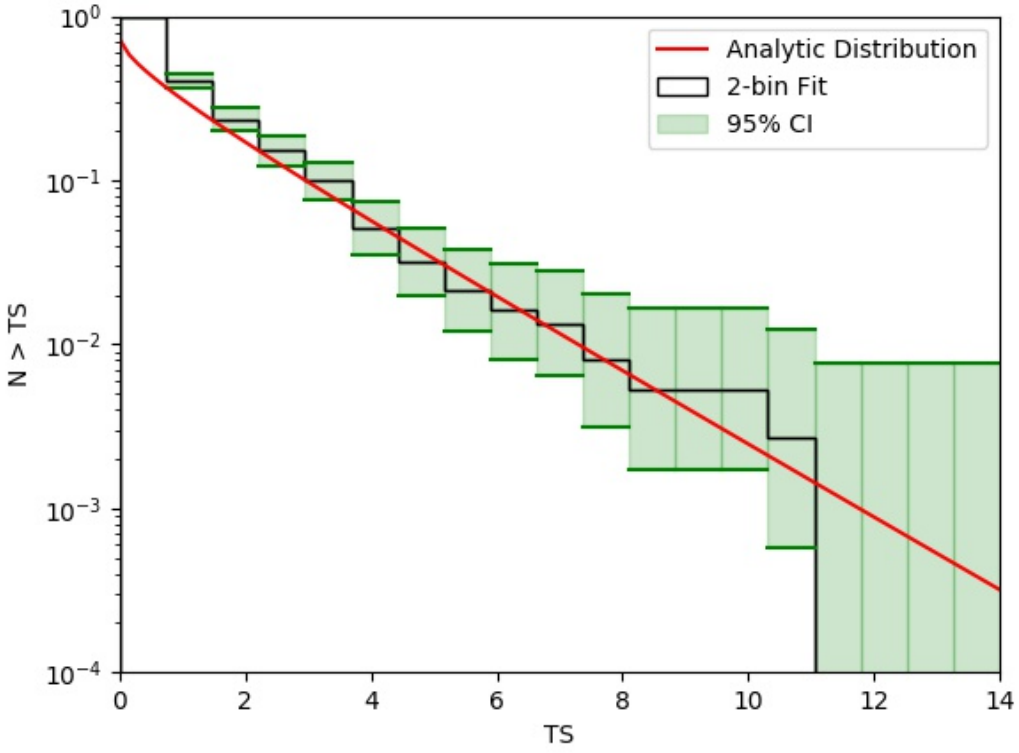
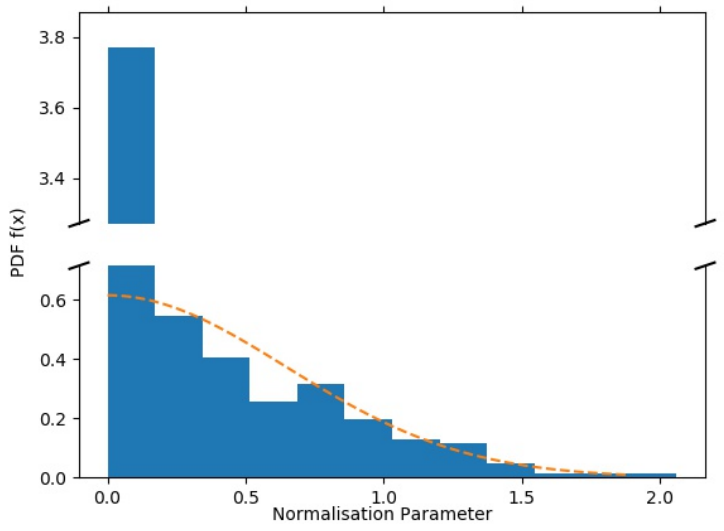
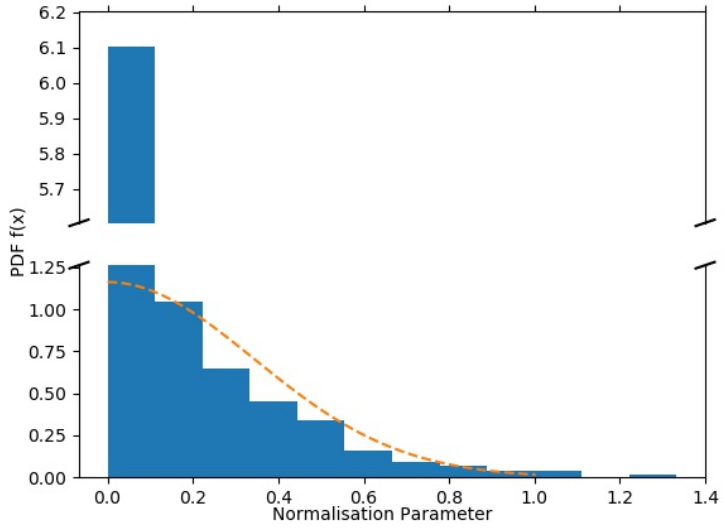


FIGURE 7.1: The empirical TS survival function distribution for a 2-bin multiband analysis. A green 95% confidence interval region is shown around the histogram data, along with the analytical function shown in red.



(a) Band 1



(b) Band 2

FIGURE 7.2: The expected distribution of the normalisation parameter of the alternative hypothesis is shown in an orange dashed line. The results of the Monte Carlo simulations for the 2 bin multiband analysis are shown as the blue histogram. The distribution parameter, σ_N^2, is estimated as $\sigma_N^2 = 0.4196$ in band 1, and $\sigma_N^2 = 0.1179$ in band 2.

7.3.2 3 Bins

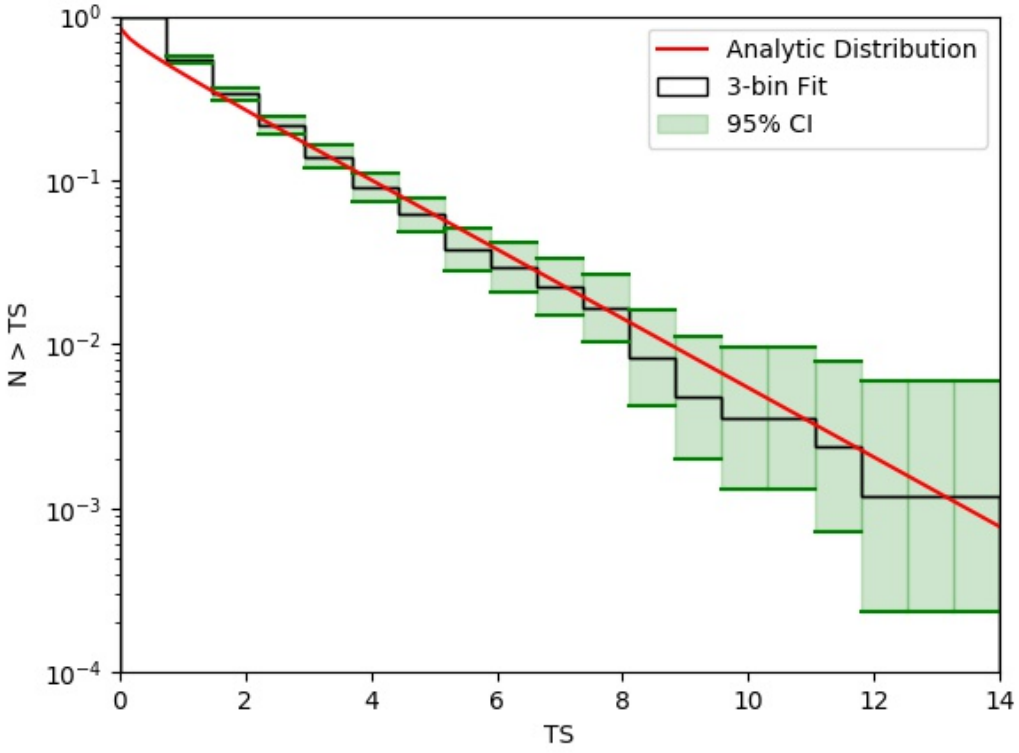
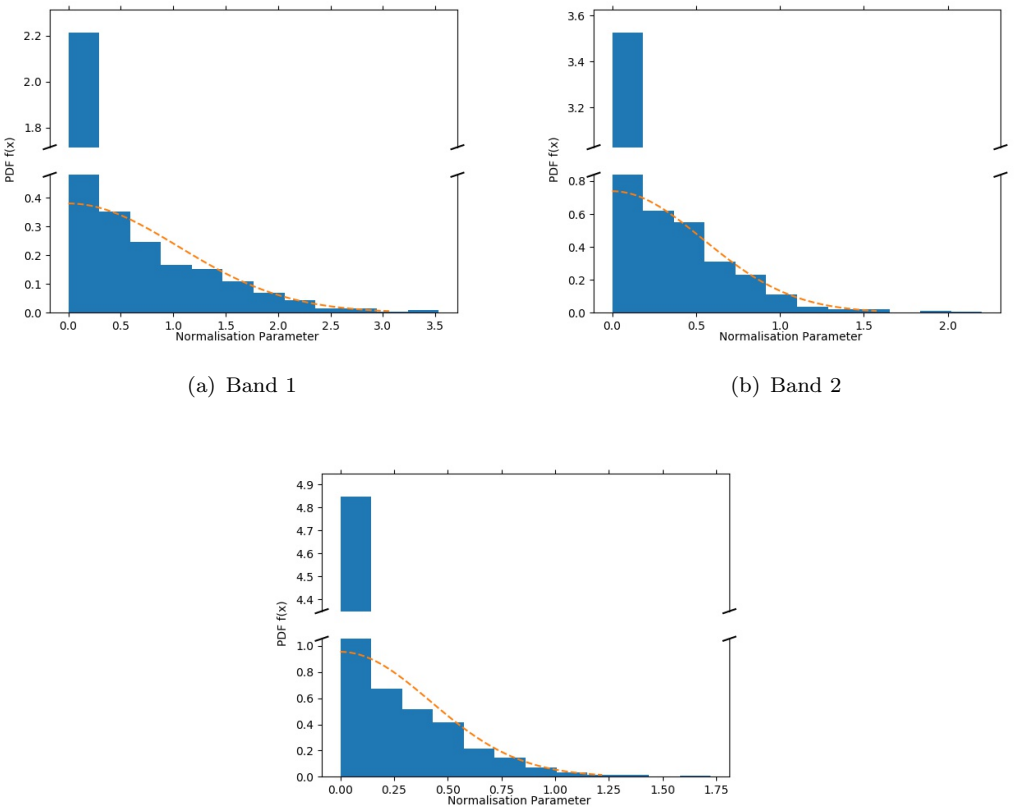


FIGURE 7.3: The empirical TS survival function distribution for a 3-bin multiband analysis. A green 95% confidence interval region is shown around the histogram data, along with the analytical function shown in red.



(a) Band 1

(b) Band 2

(c) Band 3

FIGURE 7.4: The expected distribution of the normalisation parameter of the alternative hypothesis is shown in an orange dashed line. The results of the Monte Carlo simulations for the 3 bin multiband analysis are shown as the blue histogram. The distribution parameter, σ_N^2, is estimated as $\sigma_N^2 = 1.0983$ in band 1, $\sigma_N^2 = 0.2916$ in band 2, and $\sigma_N^2 = 0.1744$ in band 3.

7.3.3 4 Bins

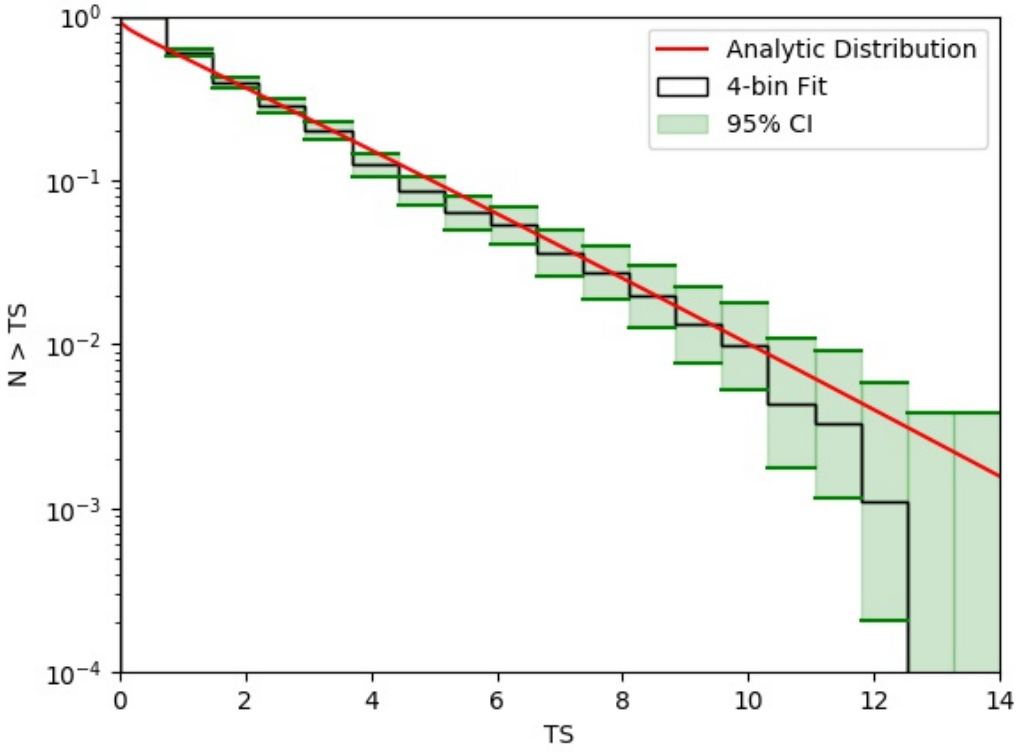
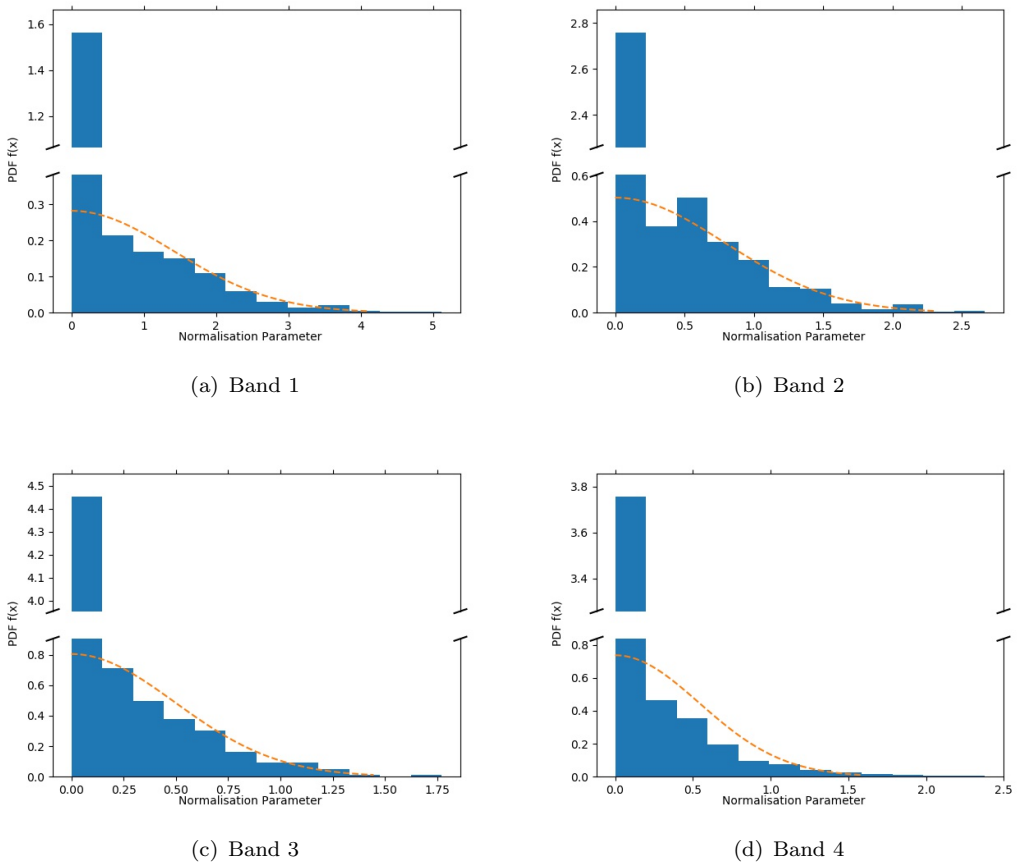


FIGURE 7.5: The empirical TS survival function distribution for a 4-bin multiband analysis. A green 95% confidence interval region is shown around the histogram data, along with the analytical function shown in red.



(a) Band 1

(b) Band 2

(c) Band 3

(d) Band 4

FIGURE 7.6: The expected distribution of the normalisation parameter of the alternative hypothesis is shown in an orange dashed line. The results of the Monte Carlo simulations for the 4 bin multiband analysis are shown as the blue histogram. The distribution parameter, σ_N^2, is estimated as $\sigma_N^2 = 1.9930$ in band 1, $\sigma_N^2 = 0.6289$ in band 2, $\sigma_N^2 = 0.2456$ in band 3, and $\sigma_N^2 = 0.2933$ in band 4.

7.3.4 5 Bins

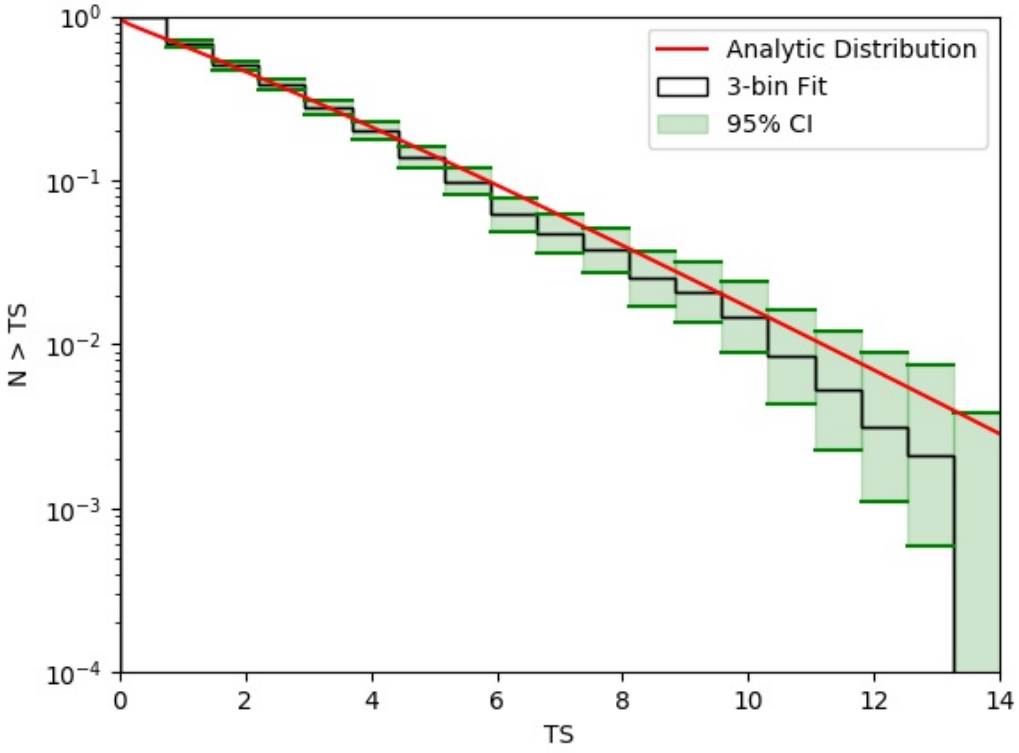
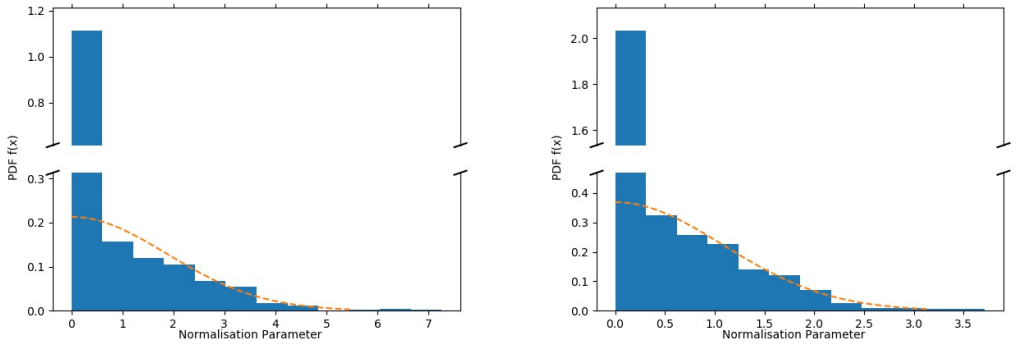
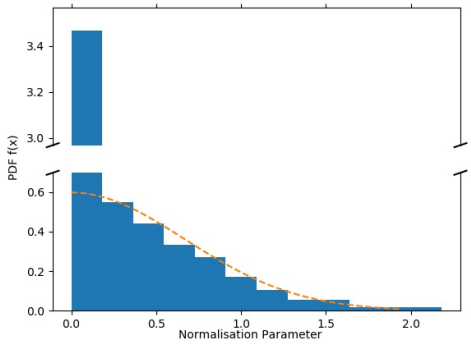


FIGURE 7.7: The empirical TS survival function distribution for a 5-bin multiband analysis. A green 95% confidence interval region is shown around the histogram data, along with the analytical function shown in red.

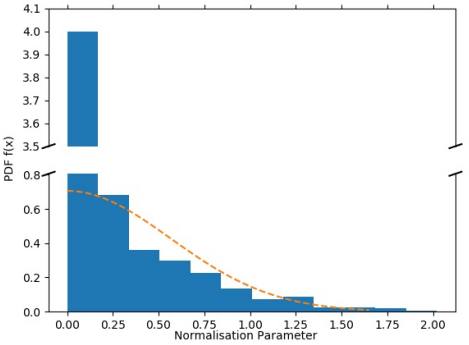


(a) Band 1

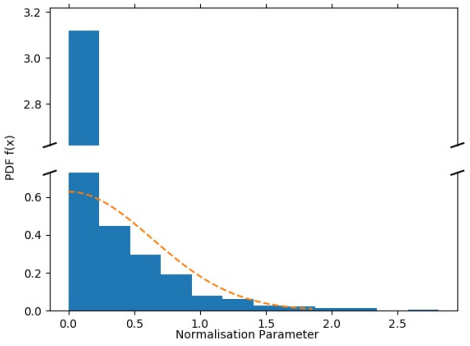
(b) Band 2



(c) Band 3



(d) Band 4



(e) Band 5

FIGURE 7.8: The expected distribution of the normalisation parameter of the alternative hypothesis is shown in an orange dashed line. The results of the Monte Carlo simulations for the 4 bin multiband analysis are shown as the blue histogram. The distribution parameter, σ_N^2, is estimated as $\sigma_N^2 = 3.5022$ in band 1, $\sigma_N^2 = 1.1701$ in band 2, $\sigma_N^2 = 0.4449$ in band 3, $\sigma_N^2 = 0.3197$ in band 4, and $\sigma_N^2 = 0.4025$ in band 5.

Chapter 8

Conclusion

The Fermi Large Area Telescope has provided an unprecedented view of the gamma ray sky, providing nine years worth of data to be analysed. Though Fermi has not yet come to the end of its mission lifespan, the data acquired has provided valuable insights into the various populations of sources, the origin of diffuse emissions, astrophysical particle accelerators, and even the possibility of a dark matter gamma ray signal.

Chapter Two introduced the field of gamma ray astronomy and detailed the means of gamma ray production, the various gamma ray sources, and their subsequent interactions with matter. These interactions motivated a number of design principles that constrained the development of pair-production gamma ray telescopes. The Fermi LAT subsystems were described and how a gamma ray signal is detected and reconstructed.

Direct measurements of a particular source were shown to be largely unattainable due to the large angular resolution of the Fermi telescope and the relative proximity of sources. Instead, Chapter Three introduced the counts quantity, which was derived from a model of various sources particle flux density, that could be compared between data and model. The particle flux density was folded through the Instrument Response Functions of the LAT to account for the uncertainties and dispersion effects of reconstructing a gamma ray event.

The maximum likelihood method of model fitting was then presented to describe the way in which to find the choice of parameters that best fit the model to the data. This was achieved by finding the parameters that maximised the probability of observing the data given the particular model. The detection of gamma ray emissions is assumed to follow a Poissonian distribution, and so the likelihood function was derived for this case. This method has been utilised in previous photon counting experiments, notably EGRET.

In order to achieve the maximum likelihood method practically, the FermiTools suite of functions were developed for use with Fermi data. These functions carry out common tasks in typical analyses, while maintaing the freedom to apply them as the particular analysis requires. Chapter Five described the `gtobssim` tool for simulating gamma ray data, and how various other tools cut and binned this data to be used by `gtlike`. The algorithms in `gtlike` are able to fit a particular model to the data, in order to maximise the likelihood function.

When two competing models are offered as potential explanations of a data set, a hypothesis test can be carried out to determine which of the models is the better choice. A test statistic is calculated for the data and compared to a threshold value, in order to reject or fail to reject one of the hypotheses. The Neyman-Pearson lemma states that the most powerful test statistic is the likelihood ratio. The likelihood ratio was then shown to be asymptotically distributed as a χ^2 distribution by Wilks theorem.

In cases such as the detection of point sources against a background, the true parameter of the underlying model may lie on the boundary of parameter space. If the background source does not have a point source present, then the normalisation parameter is zero and is the minimum possible value. Wilks theorem no longer holds for this situation. However, it was presented and derived that in a multiband analysis the likelihood ratio follows a mixture distribution of weighted χ^2 terms.

This mixture distribution was then empirically shown through the use of Monte Carlo simulations. 1000 simulations of an isotropic background were fit both by a model of just an isotropic background, and with a model of the background and a point source present. The results show empirically that the asymptotic distribution of the multiband fitting that was derived is valid, as the histogram results clearly followed the trend of the expected distribution.

The search for a dark matter gamma ray signal in Fermi data is an ongoing effort. The Galactic Center is thought to be a prime region to study this signal, however there are significant uncertainties in model parameters due to gas fields and a busy arrangement of point sources. Detailed studies of this region using a multiband analysis may help provide more fruitful evidence of the dark matter signal, because of the natural advantages of the multiband analysis. It is hoped that the distributions derived and validated in this study can be used in analyses such as this in the near future.

Bibliography

[1] Johannes Blümer, Ralph Engel, and Jörg R Hörandel. Cosmic rays from the knee to the highest energies. *Progress in Particle and Nuclear Physics*, 63(2):293–338, 2009.

[2] Volker Beckmann and Chris R Shrader. The AGN phenomenon: open issues. *arXiv preprint arXiv:1302.1397*, 2013.

[3] AA Abdo, Markus Ackermann, Marco Ajello, WB Atwood, Luca Baldini, J Ballet, Guido Barbiellini, D Bastieri, BM Baughman, K Bechtol, et al. Fermi large area telescope observations of the cosmic-ray induced γ-ray emission of the earths atmosphere. *Physical Review D*, 80(12):122004, 2009.

[4] W al M Yao et al. Review of particle physics. *Journal of Physics G: Nuclear and Particle Physics*, 33(1):1, 2006.

[5] Markus Ackermann, M Ajello, A Albert, A Allafort, L Baldini, G Barbiellini, D Bastieri, K Bechtol, R Bellazzini, E Bissaldi, et al. Search for gamma-ray spectral lines with the fermi large area telescope and dark matter implications. *Physical Review D*, 88(8):082002, 2013.

[6] Markus Ackermann, Marco Ajello, A Albert, A Allafort, WB Atwood, Magnus Axelsson, Luca Baldini, J Ballet, Guido Barbiellini, D Bastieri, et al. The fermi large area telescope on orbit: event classification, instrument response functions, and calibration. *The Astrophysical Journal Supplement Series*, 203(1):4, 2012.

[7] RC Hartman, DL Bertsch, SD Bloom, AW Chen, P Deines-Jones, JA Esposito, CE Fichtel, DP Friedlander, SD Hunter, LM McDonald, et al. The third EGRET catalog of high-energy gamma-ray sources. *The Astrophysical Journal Supplement Series*, 123(1):79, 1999.

[8] Fabio Acero, Markus Ackermann, M Ajello, A Albert, WB Atwood, Magnus Axelsson, L Baldini, J Ballet, G Barbiellini, D Bastieri, et al. Fermi large area telescope third source catalog. *The Astrophysical Journal Supplement Series*, 218(2):23, 2015.

[9] Cicerone: Effective Area. fermi.gsfc.nasa.gov/ssc/data/analysis/documentation/Cicerone/Cicerone_LAT_IRFs/IRF_EA.html. Accessed: 6th Sep 2016.

[10] Cicerone: The Point Spread Function. fermi.gsfc.nasa.gov/ssc/data/analysis/documentation/Cicerone/Cicerone_LAT_IRFs/IRF_PSF.html. Accessed: 6th Sep 2016.

[11] Pass 8 Analysis and Energy Dispersion. fermi.gsfc.nasa.gov/ssc/data/analysis/documentation/Pass8_edisp_usage.html, . Accessed: 6th Sep 2016.

[12] James R Mattox, DL Bertsch, J Chiang, BL Dingus, SW Digel, JA Esposito, JM Fierro, RC Hartman, SD Hunter, G Kanbach, et al. The likelihood analysis of egret data. *The Astrophysical Journal*, 461:396, 1996.

[13] Leif Gerward. Paul Villard and his discovery of gamma rays. *Physics in Perspective (PIP)*, 1(4):367–383, 1999.

[14] CE Fichtel, RC Hartman, DA Kniffen, DJ Thompson, GF Bignami, H Ögelman, MF Özel, and T Tümer. High energy gamma ray results from the second small astronomy satellite. *Astrophysical Journal*, 198:163–182, 1974.

[15] G F Bignami. The COS-B experiment for gamma-ray astronomy. *Space Science Instrumentation*, 1:245–268, 1975.

[16] DJ Thompson, DL Bertsch, CE Fichtel, RC Hartman, R Hofstadter, EB Hughes, SD Hunter, BW Hughlock, G Kanbach, DA Kniffen, et al. Calibration of the energetic gamma-ray experiment telescope (EGRET) for the Compton gamma-ray observatory. *The Astrophysical Journal supplement series*, 86:629–656, 1993.

[17] Malcolm S Longair. *High energy astrophysics*. Cambridge university press, 2011.

[18] Gianfranco Bertone, Dan Hooper, and Joseph Silk. Particle dark matter: Evidence, candidates and constraints. *Physics Reports*, 405(5):279–390, 2005.

[19] Chris Gordon and Oscar Macias. Dark matter and pulsar model constraints from galactic center fermi-lat gamma-ray observations. *Physical Review D*, 88(8):083521, 2013.

[20] Francesca Calore, Ilias Cholis, Christopher McCabe, and Christoph Weniger. A tale of tails: dark matter interpretations of the fermi gev excess in light of background model systematics. *Physical Review D*, 91(6):063003, 2015.

[21] F A Aharonian and A M Atoyan. Broad-band diffuse gamma ray emission of the galactic disk. *Astron. Astrophys*, 362:937–952, 2000.

[22] Enrico Fermi. On the origin of the cosmic radiation. *Physical Review*, 75(8):1169, 1949.

[23] WI Axford. Acceleration of cosmic rays by shock waves. In *Invited Papers*, pages 155–203. Springer, 1969.

[24] AA Abdo, Markus Ackermann, Marco Ajello, WB Atwood, Luca Baldini, J Ballet, Guido Barbiellini, D Bastieri, BM Baughman, K Bechtol, et al. Spectrum of the isotropic diffuse gamma-ray emission derived from first-year fermi large area telescope data. *Physical Review Letters*, 104(10):101101, 2010.

[25] Marek Sikora, Mitchell C Begelman, and Martin J Rees. Comptonization of diffuse ambient radiation by a relativistic jet: The source of gamma rays from blazars? *The Astrophysical Journal*, 421:153–162, 1994.

[26] V Beckmann and Shrader C. *Active Galactic Nuclei*. Wiley, 2012. ISBN 978-3-527-41078-1.

[27] Antony Hewish, S Jocelyn Bell, JDH Pilkington, Paul Frederick Scott, and Robin Ashley Collins. Observation of a rapidly pulsating radio source. *Nature*, 217(5130):709–713, 1968.

[28] AA Abdo, Markus Ackermann, Marco Ajello, WB Atwood, M Axelsson, Luca Baldini, J Ballet, Guido Barbiellini, MG Baring, D Bastieri, et al. A population of gamma-ray millisecond pulsars seen with the fermi large area telescope. *Science*, 325(5942):848–852, 2009.

[29] Peter Goldreich and William H Julian. Pulsar electrodynamics. *The Astrophysical Journal*, 157:869, 1969.

[30] Joseph K Daugherty and Alice K Harding. Polar cap models of gamma-ray pulsars: Emision from single poles of nearly aligned rotators. *The Astrophysical Journal*, 429:325–330, 1994.

[31] James Chiang and Roger W Romani. An outer gap model of high-energy emission from rotation-powered pulsars. *arXiv preprint astro-ph/9401034*, 1994.

[32] Alex G Muslimov and Alice K Harding. High-altitude particle acceleration and radiation in pulsar slot gaps. *The Astrophysical Journal*, 606(2):1143, 2004.

[33] Markus Ackermann, Marco Ajello, WB Atwood, Luca Baldini, J Ballet, Guido Barbiellini, D Bastieri, K Bechtol, R Bellazzini, B Berenji, et al. Fermi-lat observations of the diffuse γ-ray emission: Implications for cosmic rays and the interstellar medium. *The Astrophysical Journal*, 750(1):3, 2012.

[34] AAm Abdo, Markus Ackermann, Marco Ajello, Luca Baldini, J Ballet, Guido Barbiellini, D Bastieri, K Bechtol, R Bellazzini, B Berenji, et al. Fermi large area telescope observations of two gamma-ray emission components from the quiescent sun. *The Astrophysical Journal*, 734(2):116, 2011.

[35] Markus Ackermann, Marco Ajello, A Albert, A Allafort, Luca Baldini, Guido Barbiellini, D Bastieri, K Bechtol, R Bellazzini, E Bissaldi, et al. High-energy gamma-ray emission from solar flares: Summary of fermi large area telescope detections and analysis of two m-class flares. *The Astrophysical Journal*, 787(1):15, 2014.

[36] Jay P Norris. Fermi observations of γ-ray emission from the moon. *The Astrophysical Journal*, 2012.

[37] Felix Aharonian, AG Akhperjanian, AR Bazer-Bachi, M Beilicke, W Benbow, D Berge, K Bernlöhr, C Boisson, O Bolz, V Borrel, et al. The hess survey of the inner galaxy in very high energy gamma rays. *The Astrophysical Journal*, 636 (2):777, 2006.

[38] WB Atwood, Aous A Abdo, Markus Ackermann, W Althouse, B Anderson, M Axelsson, L Baldini, J Ballet, DL Band, Guido Barbiellini, et al. The large area telescope on the fermi gamma-ray space telescope mission. *The Astrophysical Journal*, 697(2):1071, 2009.

[39] Paolo Padovani. The blazar sequence: validity and predictions. *Astrophysics and Space Science*, 309(1-4):63, 2007.

[40] Charles D Dermer. Statistics of cosmological black hole jet sources: blazar predictions for the gamma-ray large area space telescope. *The Astrophysical Journal*, 659 (2):958, 2007.

[41] Marco Pierbattista. Gamma-ray pulsar physics: gap-model populations and light-curve analyses in the fermi era. *arXiv preprint arXiv:1309.5982*, 2013.

[42] Lars Bergström. Non-baryonic dark matter: observational evidence and detection methods. *Reports on Progress in Physics*, 63(5):793, 2000.

[43] Ivan R King. The structure of star clusters. III. some simple dynamical models. *The Astronomical Journal*, 71:64, 1966.

[44] Bing Chen Frank Haberl Leo Metcalfe Andrew M.T. Pollock Andrew M. Read Richard D. Saxton Steven Sembay Michael J.S. Smith Marcus G. F. Kirsch, Bruno Altieri. XMM-newton (cross)-calibration. *Proc.SPIE*, 5488:5488 – 5488 – 12, 2004.

[45] Cicerone: Energy Dispersion. fermi.gsfc.nasa.gov/ssc/data/analysis/documentation/Cicerone/Cicerone_LAT_IRFs/IRF_E_dispersion.html, . Accessed: 6th Sep 2016.

[46] MICHAEL Lampton, BRUCE Margon, and STUART Bowyer. Parameter estimation in X-ray astronomy. *The Astrophysical Journal*, 208:177–190, 1976.

[47] Webster Cash. Parameter estimation in astronomy through application of the likelihood ratio. *The Astrophysical Journal*, 228:939–947, 1979.

[48] J Neyman and ES Pearson. On the problem of the most efficient tests of statistical hypotheses. *J. Neyman and ES Pearson*, pages 140–85, 1967.

[49] Samuel S Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1):60–62, 1938.

[50] Herman Chernoff. On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics*, pages 573–578, 1954.

[51] Steven G Self and Kung-Yee Liang. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398):605–610, 1987.

[52] Glen Cowan. *Statistical data analysis*. Oxford University Press, 1998.